Three Studies in Synthetic Biology

Luke Cartlidge

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of MSc by Research in the Faculty of Life Sciences

Word count: 17,449

Abstract:

This report includes three separate studies in different areas of synthetic biology.

The first study's aim was to investigate the potential use of Transcription Activator-Like Effectors (TALEs) as custom-built artificial transcription factors for use in synthetic gene regulatory networks in prokaryotes. A TALE was assembled from subunits to bind the *lac* operator sequence. This was expressed from an inducible expression plasmid in *E.coli* and was able to repress expression of green fluorescent protein from a reporter plasmid.

The second study aimed to design a synthetic gene regulatory network split between two separate microbial strains which could regulate the population densities of both strains. The design was tested *in silico* and its functioning and parameter ranges were modelled.

The third study was concerned with the goal of manipulating meiotic recombination in plants, by using the CRISPR/Cas9 gene editing system to create breaks in aligned homologous chromosomes during meiosis. A test system in *Arabidopsis* was devised. A sgRNA:Cas9 construct was assembled and replication of a published edit was attempted. A meiocyte specific construct was also assembled, but not tested.

I would like to thank the following for their support and guidance: Keith Edwards, Claire Grierson, Mark Winfield, Mario Di Bernardo, Gianfranco Fiore, Fabio Annunziata, Nigel Savery and Abby Smith.

Table of Contents

Figures and Tables

Study 1

Study 2

Study 3

Transcriptional Regulation in Bacteria using TALEs

ABSTRACT

When building synthetic gene regulatory networks, the complexity that can be achieved is limited by the number of different transcription factors (TFs) available. To build truly complex circuits we would require many TFs with binding sequences orthogonal to one another and also to the chassis' native TFs. Ideally these TFs would have a range of strengths to allow tuneable control of gene expression. Transcription activator-like effectors (TALEs) are modular proteins that can be designed to predictably bind any desired DNA sequence. TALEs can therefore be exploited to create synthetic transcription factors.

We assembled a TALE, using a custom TALE kit, specific to the *lac* operator sequence. We created an expression plasmid with our TALE expressed from the inducible *araBAD* (P_{BAD}) promoter, and a reporter plasmid with the *lac* operator sequence between the *lacUV5* promoter and a *sfGFP* reporter gene.

Cells containing the TALE expression plasmid had up to a 50-fold reduction in fluorescence compared to cells with the reporter plasmid only. This reduction was similar to that achieved with the wild type LacI repressor protein expressed from the same plasmid. The reduction increased with increasing concentrations of arabinose, which induces the P_{BAD} promoter. The results strongly suggest that our TALE successfully binds the *lac* operator sequence, repressing expression of GFP. Our long-term goal is to produce a library of orthogonal, synthetic repressor and activator TFs, built *de novo* utilising TALEs as the DNA binding domain and pairs of alpha helical coiled-coils to mediate protein-protein interactions.

BACKGROUND

Synthetic Transcription Factors

One aim of synthetic biology is to build gene regulatory networks, analogous to electrical circuits, to perform desired functions and to explore the properties of natural networks (1). Networks have been built to perform a range of functions including logic operations (2), pattern formation (3), edge detection (4) and cancer cell detection (5) to name just a few. The regulation in synthetic gene networks is often applied at the transcriptional level, using transcription factors (TFs) (1). The complexity of the networks that can be built is therefore limited by the number of independent TFs available (6). As an analogy, it is like building an electrical circuit with only a handful of switches or transistors. To confound matters, existing TFs are often native to the chassis cell and so have unwanted interference with the cells' endogenous gene regulation (7). Only half a dozen or so transcription factors are regularly used in synthetic biology projects, even for the well-characterised *Escherichia coli*. This starkly contrasts with the hundreds of TFs involved in the cell's endogenous regulation (8). In order to build truly complex gene networks, we would need to be able to independently control the expression of many genes in parallel (9). Ideally, we would have libraries of hundreds of transcription factors (TFs), and other regulatory devices, with a range of activity strengths, all orthogonal to the chassis' regulatory proteins and to one another.

Three possible approaches to increase the number of available TFs would be: i) appropriate more TFs from nature. This has regularly been the source of new TFs, for example TetR or LuxR (8). Limitations of this approach are the work required to isolate and characterise the TFs, and compatibility issues with different chassis. ii) Well

characterised TFs can be modified through mutagenesis and directed evolution to alter their DNA recognition sequence or signal molecule (10). Again, this approach might involve considerable effort screening mutants, and there are likely to be limitations in how much proteins can be altered. iii) Create synthetic TFs. This option has the advantage that TFs could be designed to bind synthesised DNA sequences not used in natural gene regulation (11).

Synthetic TFs have been constructed by several synthetic biology research groups (6, 7, 11-17). A synthetic TF crucially needs a method of binding DNA, and for activators, a transcriptional activation domain. Zinc-fingers and more recently TALEs and CRISPR/Cas9 systems have been utilised to specifically bind a desired DNA sequence (8, 9). Transcription Activator-Like Effectors (TALEs) are an attractive means to develop artificial TFs, because they are modular proteins that can be assembled to bind potentially any specified DNA sequence (13). A species-compatible activation domain can be fused to the TALE. These characteristics have led to TALEs being rapidly adopted for range of research purposes, and kits have been developed for the assembly of custom TALEs, by several research groups (11, 12, 18).

TALEs

TALEs are a group of proteins produced by proteobacteria of the genera *Xanthomonas* and *Ralstonia*. TALEs were originally discovered in *Xanthomonas,* many species of which are plant pathogens and cause a range of economically important crop diseases (19). It was serendipitous that research into the mechanism of *Xanthomonas* infection led to the discovery of TALEs with their potential for application in molecular biology. During infection, TALEs are injected into the plant cell via needle like structures of the type 3 secretion system. By binding to the plant's DNA and activating certain genes, TALEs manipulate the host cell to the benefit of the bacteria (19).

Wild type TALEs consist of an N-terminal region containing a type III secretion signal; a central modular DNA-binding domain, referred to as the central tandem repeat (CTR) domain or central repeat domain; and a C-terminal region which includes nuclear localisation signals and an acidic transcriptional activation domain (19). The CTR domain consists of multiple consecutive repeats of a conserved 34 residue sequence. The number of repeats varies between TALEs but always ends with a half-repeat of around 20 residues. Residues 12 and 13 of the 34 residues, referred to as the repeat variable di-residue (RVD), differ between repeats and determine the nucleotide preferentially bound. The predictable match between the two amino acids in the RVD and the nucleotide bound was elucidated computationally and experimentally (20, 21). Crystal structures later revealed that the CTR forms a right handed super-helix that coils around double stranded DNA, with the RVDs facing inwards to interact with nucleotides in the major groove (**Figure 1**) (22, 23). Residue 12 forms hydrogen bonds with the TALE backbone that stabilise the structure; residue 13 contacts the sense DNA strand to form hydrogen bonds or van der Waals interactions with a specific nucleotide (22, 23): the RVDs HD and NN form hydrogen bonds with their nucleotide, and RVDs NI and NG form weaker van der Walls interactions (24). The CTR domain alone does not strongly bind DNA, and part of the N-terminal region adjacent to the CTR is known to contain a number of repeats which also form a super-helix. These repeats form very similar structures to the CTR repeats and bind DNA non-specifically (23, 25).

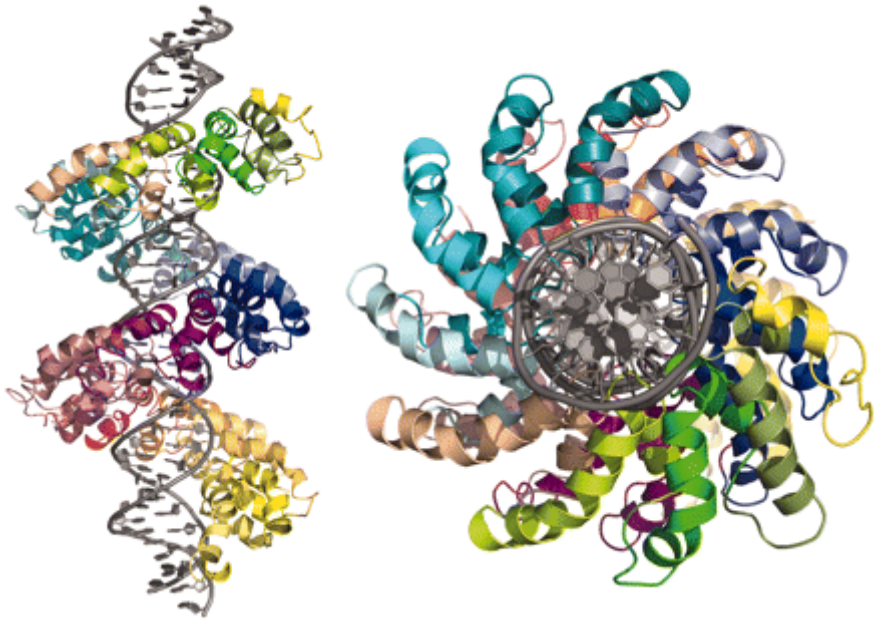

Figure 1. Structure of a TALE CTR bound to DNA, individual repeats are shown in different colours. Image taken from (23).

Their modularity and predictable specificity mean that TALEs can theoretically be designed for any desired DNA sequence. The activation domain can be replaced by other functional domains to give TALEs numerous applications (26). Artificial TALEs usually have truncated N- and C-terminals. The first 152 residues of the N-terminal region encode the secretion signal and so are unnecessary for most applications (13). The C-terminal region has been shown to be largely non-essential for DNA binding and is often truncated to 63 residues (13). The half repeat is usually included, but one study has suggested that it is an evolutionary artefact rather than a functional necessity (27). CTRs, by necessity, begin with a thymine binding RVD called the zero repeat (9). The first RVD-nucleotide correspondences deciphered were NI - A, HD - C, NG - T and NN - G (20, 21); NN will also bind A. An alternative RVD NK has higher specificity for G but lower affinity, whilst another alternative NH has improved specificity and affinity; however, most TALE designs have stuck with the original NN (9).

Artificial TALEs have been used successfully in a constantly growing list of eukaryote species, including mammals and plants (14, 15, 18, 26). TALEs have been fused to transcriptional activation domains and nucleases (15, 18). Surprisingly, there are few reports of artificial TALEs used to target genes in prokaryotes (28), but the potential to use TALEs to modify gene expression in prokaryotes has been demonstrated (6, 7, 15). Politz and colleagues successfully showed that an expressed TALE could bind to the *lac* operator and repress the expression of a downstream fluorescent reporter protein (28). Repression achieved was greater than that with the native LacI repressor protein. The TALE was designed to bind 19 base pairs of the *lac* operator O_1 sequence (29), and the DNA coding for the TALE was commercially synthesised (28). Our aim was to build on the study of Politz and colleagues by assembling TALEs to act as the DNA binding components of synthetic TFs for use in prokaryotes. The first step was to assemble a TALE from a kit and reproduce the results reported by Politz and colleagues (28). The advantage of assembling the TALE from a modular kit, rather than having the DNA synthesized, is that once assembly and assay methodologies have been established, any number of TALEs can then be assembled economically in the same

manner for orthogonal DNA sequences. Our long-term goal is to produce a library of orthogonal, synthetic repressor and activator TFs, built *de novo* utilising TALEs as the DNA binding domain and pairs of alpha helical coiled-coils to mediate protein-protein interactions, and to exploit these transcription factors for the creation of designed GRNs.

MATERIALS AND METHODS

TALE Assembly

A DNA sequence coding for a TALE complementary to 19 base pairs of the wild type *lac* operator O1 sequence (5'-TGTGAGCGGATAACAAT-3') was designed and assembled as described in (29). The assembly involved a series of hierarchical cloning steps to build a repeat array from its constituent single repeat modules (**Figure 2**). The TALE construct (**TALElac01**) was assembled from a plasmid kit available from the Joung group (Addgene kit # 1000000017). For detailed methods see the protocol (30) and original publication (18).

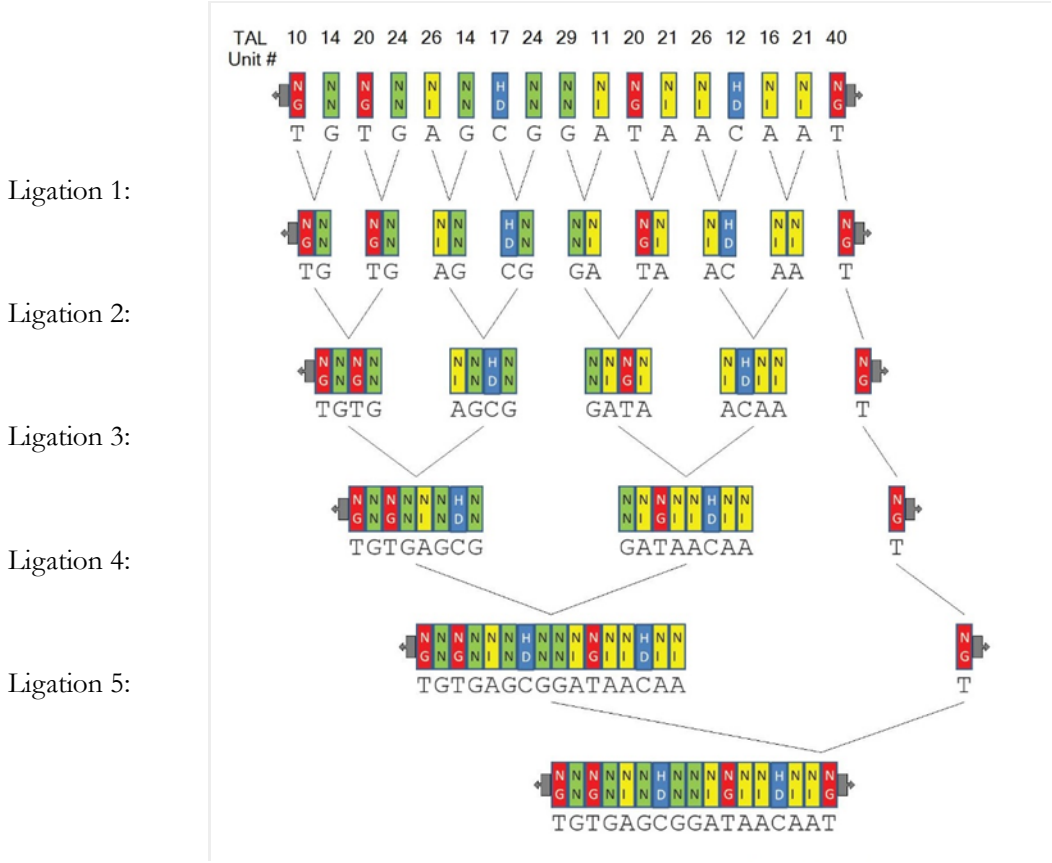

Figure 2. Schematic of the assembly of the 17-repeat array, showing the 5 hierarchical ligation steps to produce the array. A 16-repeat array minus TAL Unit 40 was cloned into the expression plasmid (before ligation 5 was carried out). The Image is taken from ZiFit software used to design the TALE (34).

Briefly: two plasmids containing the N- and C-terminal TALE repeats to be ligated were restriction digested in a 40µl total volume. The N-terminal plasmid backbone and C-terminal TALE repeat fragment were isolated by gel electrophoresis and purified. The C-terminal fragment was ligated into the N-terminal backbone. Chemically competent XL1Blue cells (Stratagene, 200249) were transformed with ligation products and grown on Luria-Bertani (LB) agar plates (Sigma-Aldrich, L7025), supplemented with 100 µg ml^{-1} carbenicillin (Thermo Fisher, 10177012), for 14 hours. 4 individual colonies were inoculated into LB broth (Sigma-Aldrich, L7275) supplemented with 100 µg ml^{-1} carbenicillin and grown for 14 hours. Plasmid DNA was purified using QIAprep Spin Miniprep Kit (QIAGEN, 27104), restriction digested and separated by gel electrophoresis to check that the insert size was correct. Five sequential cloning steps were required to assemble a 17-repeat length array, which was then cloned into an expression vector pBAD-His-JDS78 that included the N-terminal T RVD (zero repeat) and the half repeat. All

cloning steps of the assembly followed the kit protocol, with the following modifications: The digested N-terminal backbone was phosphatase treated at each step, to reduce the chance of the ends ligating — 2 µl Alkaline phosphatase (Calf intestinal) (NEB, M0290S), 5µl of 10 X Alkaline Phosphatase buffer, and 3 µl ddH_2O were mixed with the restriction digest to give 50 µl total volume and incubated at 37°C for 1 hour; polyacrylamide gels were replaced with agarose gels for electrophoresis, with gel extractions carried out using QIAEX II Gel Extraction Kit (QIAGEN, 20021); and standard T4 DNA Ligase buffer (NEB, B0202S) was used in place of quick ligase buffer.

pBAD-His-JDS78 Expression Plasmid Assembly

Our expression vector design required TALE N- and C-terminal regions downstream of the P_{BAD} promoter, to create a plasmid that would serve as a backbone for insertion of the TALE repeat array between the two terminal regions (**Table 1** shows details of all plasmids and primers used in this study).

We designed primers pBAD01 and pBAD02 for contiguous sequences including a BsmBI restriction site in the pBAD-His-RFP plasmid (Addgene, #31855), to amplify the whole plasmid. One primer included two mismatched bases to mutate out the BsmBI restriction site. The PCR mix contained: 5 µl 10 Pfu buffer, 1 µl 10 µM dNTPs, 2 µl 10 µM forward and reverse primers, 25 ng template DNA and 0.5 µl Pfu polymerase in a 50 µl final volume (reagents from Promega). Thermo-cycle conditions were: 98°C – 3 minutes, then 16 cycle of 95°C – 30 seconds, 60°C – 1 minute, 68°C – 10 minutes, followed by 68°C – 10 minutes. PCR product was mixed with 5 µl DpnI (NEB) and incubated at 37°C for 2 hours, then purified using a QIAGEN QIAquick PCR Purification Kit. 100 µl of XL1 Blue cells were mixed with 5 µl of the purified PCR product and incubated on ice for 30 minutes. The mixtures were heat shocked at 42°C for 2 minutes then returned to ice for 1 minute. 400µl of LB media was added and cells were recovered with agitation for 45 minutes at 37°C. Transformations were centrifuged at 13000rpm for 1 minute and most of the media was removed. Cells were re-suspended and 200µl of culture grown on LB plates supplemented with ampicillin 100µg ml^{-1} with agitation 300rpm, at 37 °C for 14 hours. 4 individual colonies were inoculated into 5 ml LB media supplemented with ampicillin 100 µg ml^{-1} and grown with agitation 300rpm, at 37 °C for 14 hours. Plasmid DNA was purified using a QIAGEN QIAprep Spin Miniprep Kit.

In order to add compatible restriction sites to the TALE N- + C- terminal sequence from Plasmid JDS78, a second set of primers LCTAL01.C.F and LCTAL01.C.R were designed to amplify the TALE sequence minus the FOKI domain. The primers had overhangs that included BglII and HindIII recognition sequences at the N- and C- termini respectively. The PCR mix was as above. Thermo-cycle conditions were: 98°C – 5 minutes, then 30 cycle of 94°C – 1 minute, 55°C – 1 minute, 72°C – 2 minutes, followed by 72°C – 10 minutes. The product was DpnI digested and purified as above.

The restriction-site-deleted pBAD-His-RFP plasmid and JDS78 TALE PCR product were restriction digested with NEB BglII and HindIII (NEB). The pBAD-His-RFP backbone and JDS78 TALE fragment were isolated by gel electrophoresis, ligated, transformed and purified as above. Finally the 16 repeat length array was cloned into pBAD-His-JDS78 to create plasmid **pBAD-His-TALElac01** (**Figure 3**) following the protocol from (30).

Table 1. List of Plasmids and Primers used.

Plasmid ID	Description	Insert	Source
pBAD-His-B-iRFP	Used as base for TALE/LacI expression plasmids	iRFP	Addgene
pBAD-His-JDS78	Backbone for expression of TALE repeat arrays	TALE N- and C-terminal regions	This study
pBAD-His-TALE-lac01	Expression of TALE	TALElac01 construct	This study
pBAD-His-LacI	Expression of LacI (wt)	lacI	Dr Abby Smith
TAL10 - TAL40	TALE repeat array assembly	Individual TALE repeats	(30)
JDS78	Complete TALE Assembly	TALE N- and C-terminal regions	(30)
pVRbLacUV5	Reporter plasmid for *lac* operator binding	LacUV5 SfGFP	Dr Abby Smith
Primer ID	**Sequence (5' to 3')**	**Purpose**	
LCTAL01.C.F	CGAGTCAGATCTGTGGACTTGAGGACACTCGG	Added restriction sites to JDS78 TALE sequence to clone into pBAD-His-B-iRFP	
LCTAL01.C.R	TGCGAGAAGCTTTTACGCGACTCGATGGGAAG	Added restriction sites to JDS78 TALE sequence to clone into pBAD-His-B-iRFP	
pBAD01	GCTGTGACCGATTCCGGGAGCTGC	Mutagenic PCR to delete restriction site from pBAD-His-B-iRFP	
pBAD02	GCAGCTCCCGGAATCGGTCACAGC	Mutagenic PCR to delete restriction site from pBAD-His-B-iRFP	
pBAD-LC-F	GCATGACTGGTGGACAGC	Sequencing insert in pBAD-JDS78	
pBAD-LC-R	AGACCGCTTCTGCGTTC	Sequencing insert in pBAD-JDS78	
JDS2978	TTGAGGCGCTGCTGACTG	Sequencing insert in pBAD-JDS78	
JDS2979	AAGCAATGGCGACCACCTGTTC	Sequencing insert in pBAD-JDS78	
M13 uni (-21)	TGTAAAACGACGGCCAGT	Sequencing insert in TAL10 and TAL29	
M29 rev (-29)	CAGGAAACAGCTATGACC	Sequencing insert in TAL10 and TAL29	

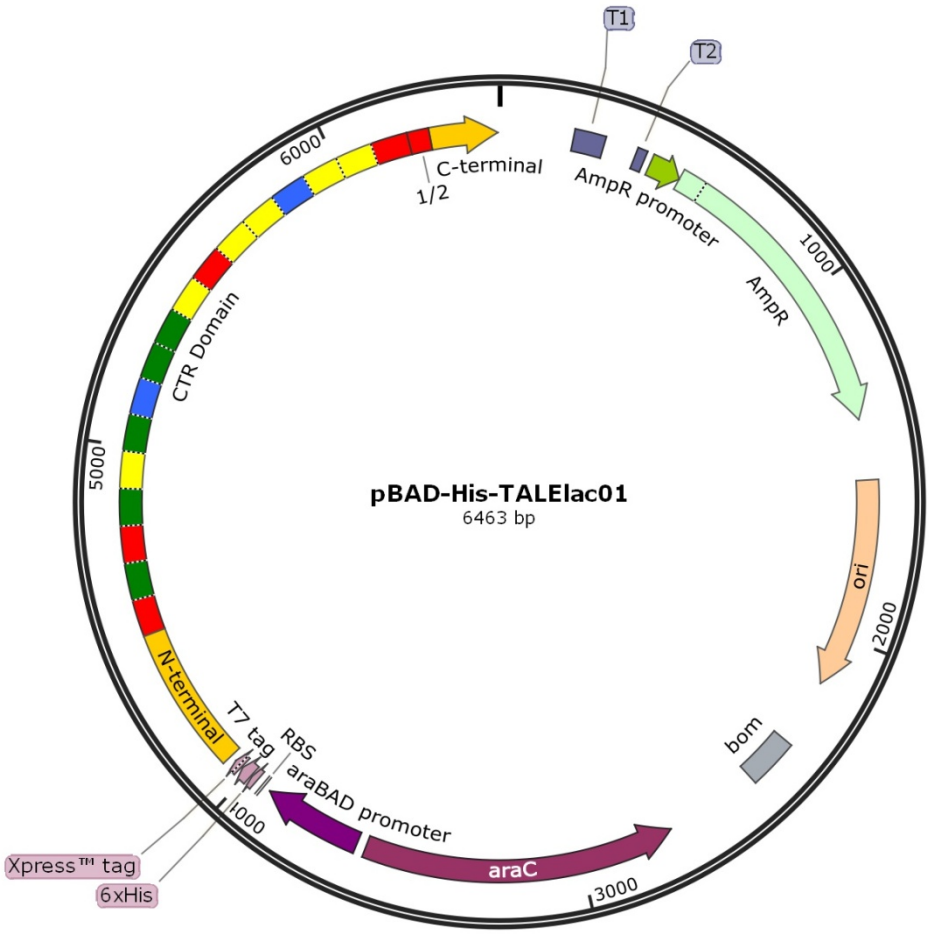

Figure 3. Plasmid map showing the final expression plasmid pBAD-His-TALElac01, including the complete TALE sequence downstream of the P_{BAD} promoter. T = terminator, AmpR = Ampicillin Resistance, CTR = Central Tandem Repeat domain of TALE, ori = origin of replication, RBS = Ribosome Binding Site, BOM = Basis of Mobility region.

Sequence Verification

Assembled TALE repeat arrays were sequence verified at the 8 and 17 repeat array stages, by PCR amplifying the array from assembly kit plasmids TAL10 or TAL29, using forward primer M13 uni (-21) and reverse primer M13 rev (-29) (See (30) for further details). The final TALElac01 insert in the **pBAD-His-TALElac01** plasmid was sequence verified from two start points using forward primers pBAD-LC-F and JDS2978 and reverse primers pBAD-LC-R and JDS2979. Sequences were aligned using GSL Biotech SnapGene Software. A combination of clear reads from the 4 pBAD-His-TALE primers gave a complete correct TALE sequence.

pVRblacUV5 and pBAD-His-LacI Plasmids

pVRblacUV5 (**Figure 4**) and **pBAD-His-LacI** (**Figure 5**) plasmids were constructed by Dr Abby Smith for a previous study. pVRblacUV5 was derived from pVRb220 (31) and had the promoter upstream of *sfGFP* replaced with the *lacUV5* promoter including the *lac* operator sequence. pBAD-His-LacI was derived from pBAD-His-RFP and had the *RFP* gene replaced with the *LacI* gene.

Fluorescence Assays

100µl of chemically competent bacterial strain TB28 (32)(Genotype: Δ*lacIZYA*) cells were mixed with 10µl, 10nM of each appropriate plasmid (**pVRblacUV5, pBAD-His-LacI, pBAD-His-TALElac01**) and grown on LB plates supplemented with the relevant antibiotic(s) (pVRblacUV5 - 50 µg ml^{-1} Kanamycin, **pBAD-His-LacI/pBAD-His-TALElac01** 100µg ml^{-1} ampicillin), for 14 hours at 37°C. Three individual colonies were inoculated into 5 ml LB media supplemented with the relevant antibiotic(s) and grown with agitation 300rpm at 37°C for 14 hours. 100µl of culture was mixed with 10ml LB media supplemented with the relevant antibiotic(s), arabinose (0.0002 to 2 % w/v) and IPTG (1mM). Cultures were then incubated at 37°C 300rpm agitation. Optical density was measured approximately every 15 minutes. Once cells had reached exponential phase, they were stored on ice. For initial optimisation experiments, cells were transferred to ice when OD600 reached approximately 0.5. For the final TALE expression assay, cells were grown for longer until OD600 reached approximately 1.4. 5 ml of each culture was centrifuged for 4 minutes at 5000rpm, 4°C. Cells were re-suspended in 200µl PBS. 100µl of the suspension was added to the well of a black Greiner 96 well plate. Fluorescence readings were taken using Molecular Devices, SpectraMax 190 Microplate Reader with the following settings: excitation wave length 485nm, Absorbance 510nm, cut-off 495nm.

Fluorimeter Data Analysis

Data analysis was performed with Microsoft Office Excel 2010. To account for background auto-fluorescence, average readings for non-transformed "empty" TB28 cells were subtracted from treatment readings. Readings were normalised by dividing by OD600 readings and shown relative to the highest reading for the replicate.

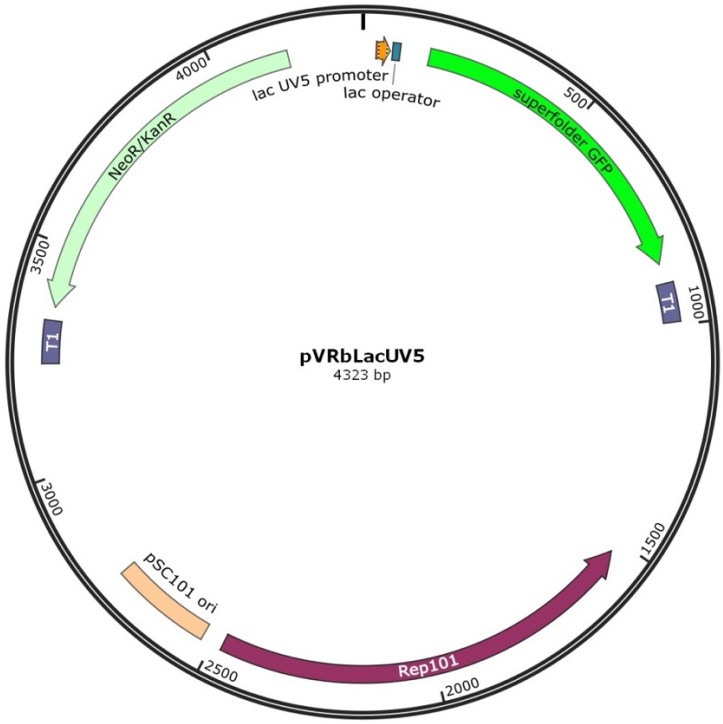

Figure 4. Plasmid map showing the reporter plasmid pVRbLacUV5, with the Lac operator O_1 sequence between the *lacUV5* promoter and *sfGFP* gene. T = terminator, KanR = Kanamycin Resistance, NeoR = Neomycin Resistance, ori = origin of replication.

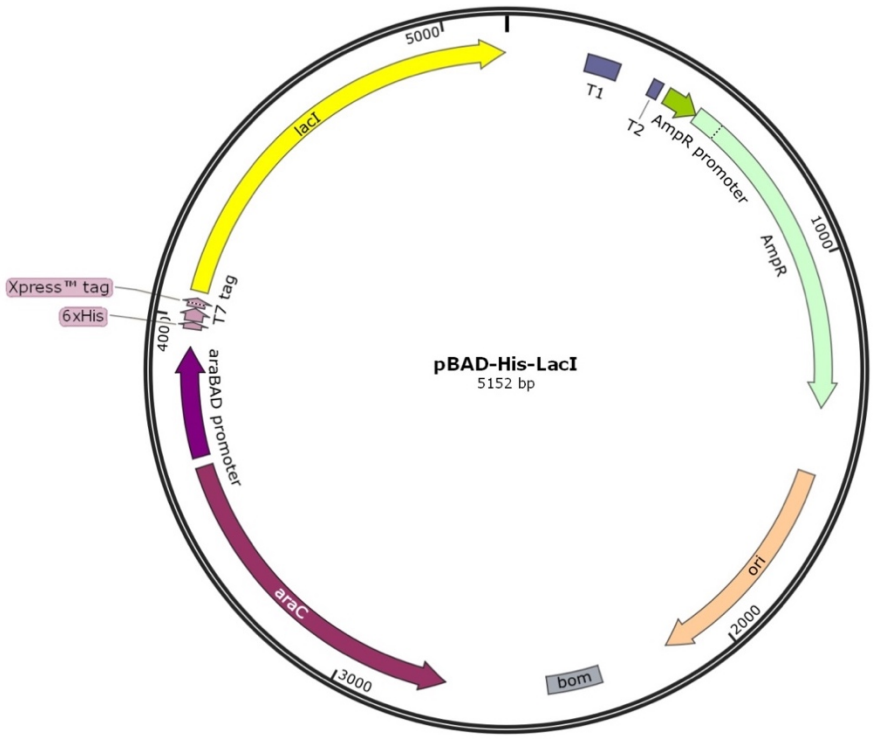

Figure 5. Plasmid map showing the expression plasmid pBAD-His-LacI, with *LacI* downstream of the P_{BAD} promoter. T = terminator, AmpR = Ampicillin Resistance, ori = origin of replication, RBS = Ribosome Binding Site, BOM = Basis of Mobility region.

RESULTS

TALE Expression Vector

We constructed a TALE expression vector — pBAD-His-JDS78 — with restriction sites between the TALE N- and C-terminal regions for the insertion of any TALE repeat array. This expression plasmid includes the standard T-binding zero repeat and a final half repeat, corresponding to T, and can be used for any designed repeat array — although the sequence must end with a T. The TALE sequence is downstream of the inducible P_{BAD} promoter and the plasmid includes the *araC* gene, so that in the presence of arabinose araC induces expression from the P_{BAD} promoter of the TALE. Therefore, expression of any TALE from the plasmid should be tightly regulated and inducible with arabinose. Aside from the change of promoter, it was also necessary to create this new expression plasmid to omit the FOKI nuclease included in the assembly kit expression vector.

TALE Repeat Array

We assembled a 17 repeat length array without N- or C-terminals (**Figure 2**). We also cloned an array of 16 of the repeats into our pBAD-His-JDS78 expression plasmid to create our final complete TALE expression plasmid — **pBAD-His-TALElac01** (**Figure 3**): the final repeat in the TALE was omitted, due to a shortage of time. This TALE sequence was verified by sequencing and could be used in future TALE constructs, as could its constituent 2, 4 and 8 repeat subunits.

TALE Binding and Repression Assays

We tested whether our TALE would bind to its target *lac* operator sequence by transforming *E. coli* cells, strain TB28 which has the *lac* repressor (*LacI)* gene deleted, with the **pBAD-His-TALElac01** expression plasmid and with a GFP reporter plasmid — **pVRbLacUV5**. **pVRbLacUV5** has the *lac* operator sequence between the *lacUV5* promoter and a green fluorescent protein (*sfGFP*) gene, therefore TALE binding at the lac operator should repress expression of GFP. The repression assays included a control with the wild type LacI repressor protein expressed from the same plasmid as our TALE — **pBAD-His-LacI**.

Prior to expression assays, we took fluorimeter readings for a range of densities of cells containing the reporter plasmid pVRbLacUV5 only, to ensure that readings had a linear relationship with fluorescence and were not becoming saturated at high cell density. Readings had a linear relationship for 5ml growth cultures concentrated by factors of 3× to 50× (**Figure 6**). Cells containing only the reporter construct were expected to give the highest fluorescence readings, and we used 25× culture concentrations in all further assays; therefore, our assayed cells were at densities well within the tested fluorescence readings range.

Before conducting the final TALE repression assay, we tested GFP expression in the presence of the **pBAD-His-LacI** plasmid at different arabinose and IPTG concentrations. This was to test whether arabinose-induced-expression from our plasmid behaved as expected, without the complication that our TALE may not function as expected. The addition of IPTG should remove LacI repression and therefore provide evidence that any decrease in GFP expression was the result of repression via the *lac* operator and not due to other factors, such as the metabolic burden introduced by the additional expression plasmid. **pBAd-His-LacI** containing cells had a twofold reduction in fluorescence compared to cells with the reporter plasmid alone (**Figure 7**). The reduction was removed partially in the presence of 1mM IPTG. However, arabinose concentration appeared to have no effect. The results indicate that the assay system overall worked as expected. The fact that repression occurs without arabinose suggests

some leakiness of the P_{BAD} promoter. One explanation for the low level of repression observed, and lack of an effect of arabinose concentration, is that more time was required for repression to become fully observable, perhaps due to the time needed to degrade accumulated GFP: cultures had only taken approximately 2.5 hours to reach 0.5 OD600. Because of this in the final assay, cells were grown for 4 hours, to an OD600 of roughly 1.4 instead of 0.5.

Finally, we tested cells transformed with our TALE expression plasmid **pBAD-His-TALElac01**. We tested the fluorescence of cells grown in arabinose concentrations of 0 - 0.2 % w/v. Without arabinose, Cells transformed with the TALE expression plasmid (**pBAD-His-TALElac01**) and reporter plasmid (**pVRbLacUV5**) had a two-fold reduction in fluorescence compared to cells transformed with the reporter plasmid alone (**Figure 8**). This reduction was similar to that achieved in cells transformed with the LacI expression plasmid (**pBAD-His-LacI**) plus the reporter plasmid. This suggests that the TALElac01 protein was expressed and bound the *lac* operator sequence as designed, blocking transcription of the *sfGFP* gene. Fluorescence decreased further with increasing arabinose concentration, with a maximum 50-fold reduction at 0.2% arabinose. The graduated reduction in fluorescence was very similar with LacI. This provides further evidence that the reduction of fluorescence was indeed the result of TALE protein expression, and shows that the level of repression could be controlled via arabinose concentration.

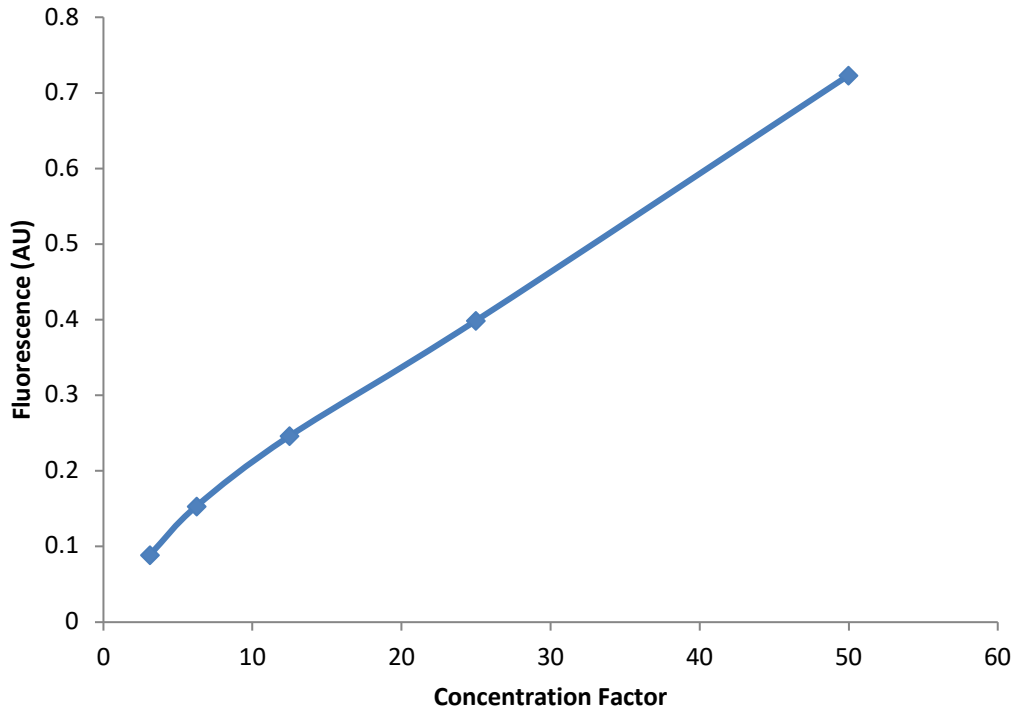

Figure 6. Fluorescence readings for different cell densities. Cells containing the sfGFP reporter plasmid (pVRbLacUV5). The linear relationship between the concentration factor of 5ml cultures and fluorescence show that the fluorimeter readings were not saturated in this range. Diamonds show means, n = 2. Fluorescence in arbitrary units normalised to range 0-1.

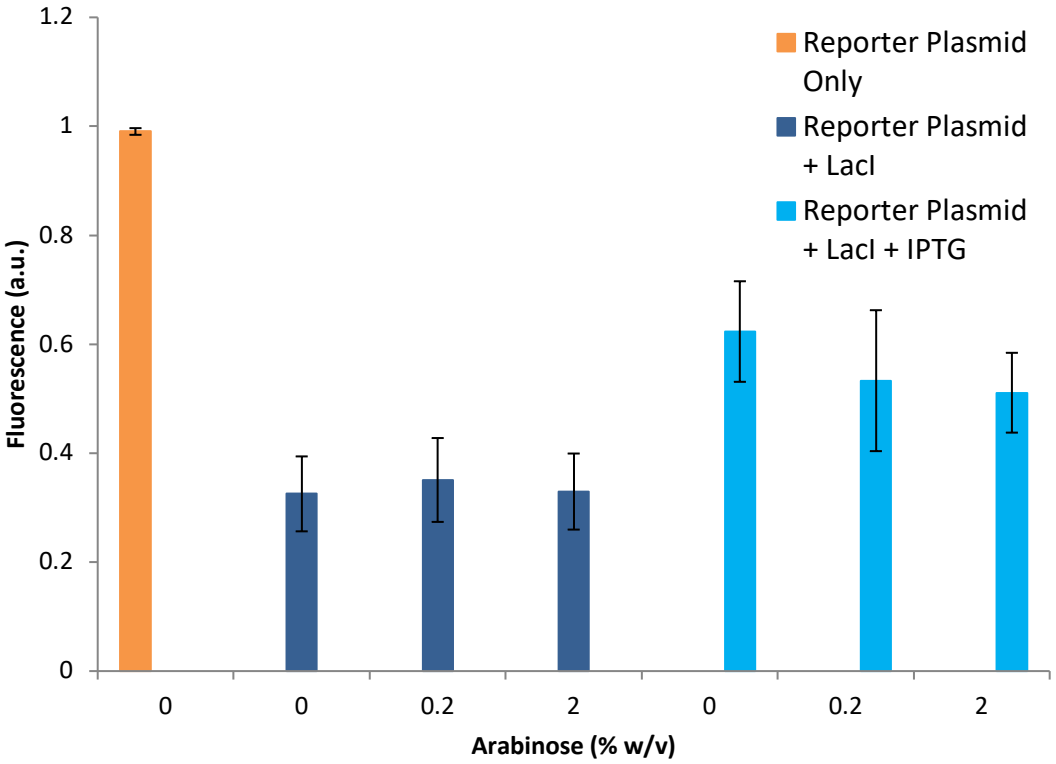

Figure 7. Fluorescence readings for cells containing GFP Reporter Plasmid Only (**pVRbLacUV5**) or **Reporter Plasmid plus LacI expression plasmid** (**pBAD-His-LacI**), with and without 1mM **IPTG**. Cultures were grown with different Arabinose concentrations. Bars show Means ± SEM, n = 3. Fluorescence in arbitrary units normalised to range 0-1.

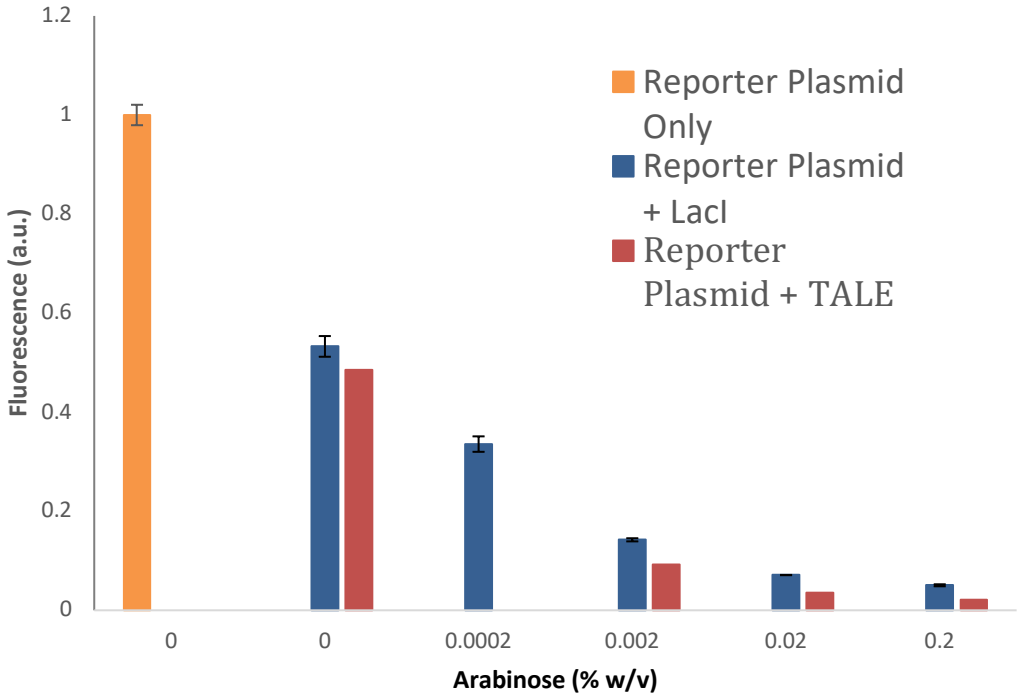

Figure 8. Fluorescence readings for cells containing **sfGFP Reporter Plasmid Only** (**pVRbLacUV5**), **Reporter Plasmid** plus **LacI expression plasmid** (**pBAD-His-LacI**) or **Reporter Plasmid** plus **TALElac01 expression plasmid** (**pBAD-His-TALElac01**). Cultures were grown with different Arabinose concentrations. Bars show Means ± SEM, n = 2 for Reporter Only and Reporter plus LacI, n = 1 for Reporter plus TALElac01. Fluorescence in arbitrary units normalised to range 0-1. No recording was made for TALE at 0.0002% Arabinose: due to an experimental error with no time available to repeat the experiment.

DISCUSSION

Our results support the findings of Politz and colleagues that TALEs can be used successfully in prokaryotes (28). From these initial results, we cannot exclude the possibility that the change in fluorescence was a direct response in GFP expression to arabinose concentration, although there is no obvious link between the two. Further controls would be needed with cells containing the reporter plasmid only grown with various arabinose concentrations. It is also possible that the reduced fluorescence was not a response to repressor protein binding, but to some other factor caused by the presence of the expression plasmid. To refute this possibility, experiments to would need to include removal of the CTR domain of the TALE and separately removal of the lac operator sequence, to demonstrate both are necessary for reduced GFP expression; however, time was not available to include these additional controls.

pBAD-His-TALElac01 assembly lays the foundation to create a range of TALES using the tested assembly method. Assembly of a 17-repeat array, using the Joung Lab kit and protocol takes 5 weeks, if everything runs smoothly. Our 17 repeat array required 5 sequential sets of cloning steps and a final step to clone the array into the expression vector. Each step requires 4-5 days, but this can easily take longer. Because the TALEs are assembled by ligating pairs of repeat arrays, 4-, 8- and 16-repeat arrays are efficient lengths to build. Our results suggest 18.5 repeats were sufficient to bind DNA, including the zero repeat and half repeat. Gene activation has been shown in plants with just 6.5 repeats and activation plateaued after 10.5 repeats (9). In future, we aim to assemble multiple TALEs in parallel using a liquid handling robot to speed up assembly.

The pBAD-His-JDS78 expression plasmid is a useful tool to assay future TALE constructs. Any repeat array assembled using the Joung group kit can be cloned directly into our plasmid which includes the TALE N- and C-terminals. The target DNA sequence needs to end with a T, because the existing plasmid contains the final half repeat RVD for T. However, it would be straightforward to construct plasmids for the other three nucleotides, using primers LCTAL01.C.F/R to amplify from other expression plasmids in the kit. Our intermediate, shorter CTR assemblies can also be utilised later to create TALEs for different target DNA sequences.

These results are a first step toward the larger goal of assembling a range of synthetic repressors and activators based on TALEs. Unpublished work by the Savery group has used *de novo* designed coiled-coil pairs (33) to mediate interaction between a *Bacteriophage lambda* cI repressor protein and the alpha subunit of RNA polymerase, creating a synthetic transcriptional activator. The coiled-coils are pairs of rationally designed alpha-helix peptides that associate through hydrophobic residues to form dimers. A set of coiled-coil pairs has been created, with a range of dissociation constants (33). The cI protein acts as the DNA binding domain of the activator. It was expressed as a fusion protein with one half of a coiled-coil pair. The second half of the coiled-coil pair is expressed as a fusion protein with the alpha subunit of RNA polymerase. The alpha-helix therefore acts as the activation domain of the synthetic activator, recruiting RNA polymerase. By placing the cI recognition DNA sequence upstream of a promoter for a reporter protein, it was shown that the synthetic TF increased gene expression in the presence of the modified alpha subunit. Activators built in this way have the same limited number of DNA binding sites as natural TFs. However, by replacing the cI protein with a TALE we hope to produce a set of orthogonal TFs in which the whole protein sequence is rationally designed for a specified DNA sequence.

The advantages of this system would be that: The enhancer sequence could be designed to avoid off target interference with the chassis species' genome. Software exists to design TALEs to avoid off target binding (15). A large number of orthogonal enhancer-activator pairs could be designed, giving parts that could be used "off the shelf" from a library, or custom built to new specifications. The strength of transcriptional activation would be tuneable via a combination of the dissociation constants of the coiled-coil pair and the DNA binding affinity of the TALE. TALE affinity can be altered by changing the number of repeats in the CTR domain or by including mismatched RVDs. TALEs are known to bind DNA sequences with up to 3 base pair mismatches (15).

Towards this goal, our next step is to create a fusion protein with TALElac01 replacing cI as the DNA binding domain of a synthetic activator (**Figure 9**). A reporter plasmid with the *lac* operator upstream of a promoter for *sfGFP* would be used to test activation strength. Another interesting extension would be to create short TALEs that only bind DNA cooperatively when dimerised via coiled-coils. Similar systems have been reported and allow a neat method to create logic gates (17).

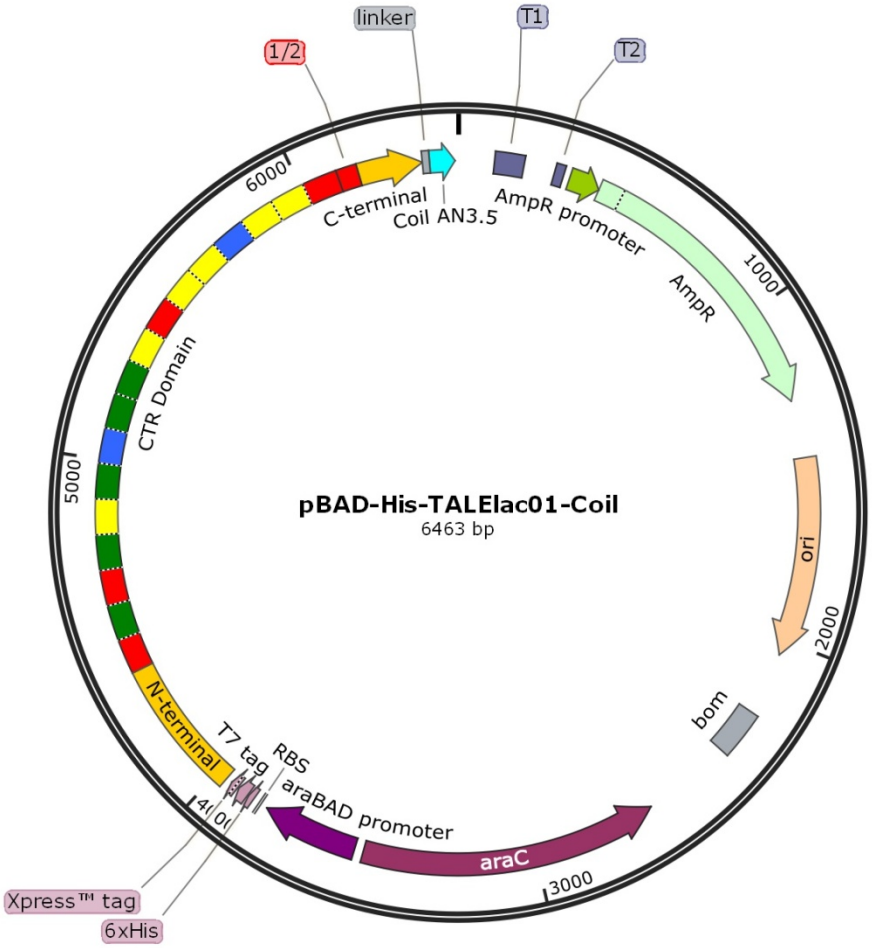

Figure 9. Plasmid map showing a design to incorporate one half of a coiled-coil pair (Coil AN3.5) into our existing pBAD-His-TALElac01 expression plasmid to create a synthetic activator. T = terminator, AmpR = Ampicillin Resistance, CTR = Central Tandem Repeat domain of TALE, ori = origin of replication, RBS = Ribosome Binding Site, BOM = Basis of Mobility region.

This system of orthogonal TFs could potentially help solve the problems of the shortage of orthogonal TFs and interference with chassis functioning. However, one limitation is that the TFs would not directly recognise chemical

signals; natural TFs are often activated or inactivated by a signal molecule. In our test system, external control of expression was still via the natural TF AraC and its ligand arabinose. The synthetic activators could be used as intermediates in regulatory networks and to create logic gates, but the networks would still rely on natural proteins to detect external signals. In electronics terms it gives us the wires but not the sensors. A solution to this problem was recently demonstrated by Rai and colleagues (7) who used a combination of TALEs and ribo-switches to create synthetic repressors only active in the presence of signal molecules. Finding other solutions to this limitation would be another interesting line of future research.

DISCLAIMER

Work All work was carried out by the student, with the exception of construction of the pBAD-His-LacI and pVRbLacUV5 plasmids, the final fluorimeter assay of TALE repression and design of primers pBAD01 and pBAD02, which were carried out by Dr Abby Smith.

REFERENCES

1. Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems Biology. 2006;2:14.
2. Wang BJ, Kitney RI, Joly N, Buck M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nature Communications. 2011;2:9.
3. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. A synthetic multicellular system for programmed pattern formation. Nature. 2005;434(7037):1130-4.
4. Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM, et al. A Synthetic Genetic Edge Detection Program. Cell. 2009;137(7):1272-81.
5. Anderson JC, Clarke EJ, Arkin AP, Voigt CA. Environmentally controlled invasion of cancer cells by engineered bacteria. Journal of Molecular Biology. 2006;355(4):619-27.
6. Li YQ, Jiang Y, Chen H, Liao WX, Li ZH, Weiss R, et al. Modular construction of mammalian gene circuits using TALE transcriptional repressors. Nature Chemical Biology. 2015;11(3):207-U156.
7. Rai N, Ferreiro A, Neckelmann A, Soon A, Yao A, Siegel J, et al. RiboTALE: A modular, inducible system for accurate gene expression control. Scientific Reports. 2015;5:11.
8. Rao CV. Expanding the synthetic biology toolbox: engineering orthogonal regulators of gene expression. Current Opinion in Biotechnology. 2012;23(5):689-94.
9. Kabadi AM, Gersbach CA. Engineering synthetic TALE and CRISPR/Cas9 transcription factors for regulating gene expression. Methods. 2014;69(2):188-97.
10. Kamionka A, Sehnal M, Scholz O, Hillen W. Independent regulation of two genes in Escherichia coli by tetracyclines and Tet repressor variants. Journal of Bacteriology. 2004;186(13):4399-401.
11. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting (vol 39, pg e82, 2011). Nucleic Acids Research. 2011;39(17):7879-.
12. Weber E, Gruetzner R, Werner S, Engler C, Marillonnet S. Assembly of Designer TAL Effectors by Golden Gate Cloning. Plos One. 2011;6(5):5.
13. Miller JC, Tan SY, Qiao GJ, Barlow KA, Wang JB, Xia DF, et al. A TALE nuclease architecture for efficient genome editing. Nature Biotechnology. 2011;29(2):143-U9.
14. Perez-Pinera P, Ousterout DG, Brunger JM, Farin AM, Glass KA, Guilak F, et al. Synergistic and Tunable Gene Activation by Combinations of Synthetic Transcription Factors. Molecular Therapy. 2013;21:S93-S.
15. Garg A, Lohmueller JJ, Silver PA, Armel TZ. Engineering synthetic TAL effectors with orthogonal target sites. Nucleic Acids Research. 2012;40(15):7584-95.
16. Farzadfard F, Perli SD, Lu TK. Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. Acs Synthetic Biology. 2013;2(10):604-13.
17. Khalil AS, Lu TK, Bashor CJ, Ramirez CL, Pyenson NC, Joung JK, et al. A Synthetic Biology Framework for Programming Eukaryotic Transcription Functions. Cell. 2012;150(3):647-58.
18. Sander JD, Cade L, Khayter C, Reyon D, Peterson RT, Joung JK, et al. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nature Biotechnology. 2011;29(8):697-8.

19. Boch J, Bonas U. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. In: VanAlfen NK, Bruening G, Leach JE, editors. Annual Review of Phytopathology, Vol 48. Annual Review of Phytopathology. 48. Palo Alto: Annual Reviews; 2010. p. 419-36.

20. Scholze H, Boch J. Breaking the code of DNA binding specificity of TAL-type III effectors. Virulence. 2010;1(5):428-32.

21. Moscou MJ, Bogdanove AJ. A Simple Cipher Governs DNA Recognition by TAL Effectors. Science. 2009;326(5959):1501-.

22. Deng D, Yan CY, Pan XJ, Mahfouz M, Wang JW, Zhu JK, et al. Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. Science. 2012;335(6069):720-3.

23. Mak ANS, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. Science. 2012;335(6069):716-9.

24. Streubel J, Blucher C, Landgraf A, Boch J. TAL effector RVD specificities and efficiencies. Nature Biotechnology. 2012;30(7):593-5.

25. Gao HS, Wu XJ, Chai JJ, Han ZF. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. Cell Research. 2012;22(12):1716-20.

26. Moore R, Chandrahas A, Bleris L. Transcription Activator-like Effectors: A Toolkit for Synthetic Biology. Acs Synthetic Biology. 2014;3(10):708-16.

27. Zheng CK, Wang CL, Zhang XP, Wang FJ, Qin TF, Zhao KJ. The last half-repeat of transcription activator-like effector (TALE) is dispensable and thereby TALE-based technology can be simplified. Molecular Plant Pathology. 2014;15(7):690-7.

28. Politz MC, Copeland MF, Pfleger BF. Artificial repressors for controlling gene expression in bacteria. Chemical Communications. 2013;49(39):4325-7.

29. Beyreuth.K. RECOGNITION OF LAC OPERATOR OF LACTOSE OPERON O E-COLI BY LAC REPRESSOR. Hoppe-Seylers Zeitschrift Fur Physiologische Chemie. 1973;354(10-1):1171-.

30. Reyon D, Khayter C, Regan MR, Joung JK, Sander JD. Protocol for REAL and REAL-Fast Assembly of TALEN Expression Plasmids 2012 [cited 2015 1 July]. Available from: http://www.talengineering.org/platforms-real.htm.

31. Rhodius VA, Segall-Shapiro TH, Sharon BD, Ghodasara A, Orlova E, Tabakh H, et al. Design of orthogonal genetic switches based on a crosstalk map of rs, anti-rs, and promoters. Molecular Systems Biology. 2013;9:13.

32. Bernhardt TG, de Boer PAJ. Screening for synthetic lethal mutants in Escherichia coli and identification of EnvC (YibP) as a periplasmic septal ring factor with murein hydrolase activity. Molecular Microbiology. 2004;52(5):1255-69.

33. Thomas F, Boyle AL, Burton AJ, Woolfson DN. A Set of de Novo Designed Parallel Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-nanomolar Regime. Journal of the American Chemical Society. 2013;135(13):5161-6.

34. Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Research. 2010;38:W462-W8.

A Population Density Regulating Synthetic Gene Network Design

ABSTRACT

The ability to control the population densities of multiple microbial populations simultaneously may be desirable for a range of synthetic biology and biotechnology applications in which microbial consortia perform more efficiently than a single strain. Many bacteria use quorum sensing based feedback systems to monitor and respond to changes in population density and this presents a means to create synthetic gene regulatory networks which regulate population density. Here we test *in silico* a design for a gene network, split between two populations, that implements negative feedback to regulate both population densities. The design was originally developed as part of project S3.2 at the Bristol Centre for Synthetic Biology and is adapted here to control a different population phenotype. Specifically, in the design, a "master" population autonomously regulates its own density, and also acts as a controller to regulate the density of the second "slave" population. Negative feedback loops in the design utilise quorum signalling pathways and the toxin protein ccdB to maintain stable densities of both populations. The steady state population densities can be tuned via an external arabinose input. We simulated the design in silico to investigate its properties and parameter constraints. Our model showed that the design is able to keep population densities steady and that densities can be modulated as required. This proof of concept design could be expanded to control multiple-strain microbial consortia.

BACKGROUND

One aim of synthetic biology is to build gene regulatory networks, analogous to electrical circuits, to perform desired functions and to explore the properties of natural networks (1). A great deal of progress has been made in engineering single cells to perform new functions, for example chemical production, logic operations, cancer cell detection and sensing (2-5). While the majority of studies have created synthetic circuits embedded in single cells, the complexity and efficiency that can be achieved may be increased by engineering multicellular systems in which functionality is distributed amongst cells (6, 7). Multicellular systems provide the following advantages: the metabolic burden placed on cells, by introducing biological components, can be shared between strains, potentially allowing more complex circuits to be assembled (8); stochasticity resulting from intrinsic variation between individual cells may be smoothed out, if signals are taken as the averages from many cells (9); and certain natural properties of a species may be impossible to introduce into another species, and therefore the only way to combine properties may be to have multiple species growing together (7).

Population Density Regulation

For many potential applications of microbial consortia, it would be necessary to regulate population density reliably, in terms of both absolute density and the ratios between individual populations (6). Engineered autonomously-regulating cells could achieve this population control and alleviate some of the need for constant external regulation of growth conditions. In fermentation or pharmaceutical production, for instance, it could be important to keep bacterial strains at set densities to prevent the build-up of toxic metabolites reducing growth, or to optimise the efficiency of a process. Autonomously regulating bacterial populations could also be applied as therapeutics (7, 10);

an engineered bacterial strain could be introduced into a patient, which would thereafter maintain its own density and release a therapeutic at the correct rate. Ideally, for a hypothetical synthetic consortium, population densities would autonomously remain steady and would be tuneable via external inputs. Multicellular systems require inter-cellular communication for coordinated behaviour (11), and in bacteria this is usually achieved by quorum sensing (QS) systems (12). Quorum sensing describes a coordinated response in bacteria to changes in population density. Quorum sensing bacteria release signalling molecules — known as autoinducers — into their surroundings, which at a threshold concentration trigger cascades of gene regulation. In general, the autoinducers of gram-negative bacteria are N-acyl homoserine lactones — (NA)HSLs — which differ in the number of carbon atoms in the side chain.

The canonical quorum sensing system is that of the bioluminescent bacterium *Aliivibrio fischeri* (formerly *Vibrio fischeri*) which lives symbiotically within the Hawaiian bobtail squid, *Euprymna scolopes* (as well as within other marine species). When the bacteria live at high densities, within the light-organ of the squid, a selective advantage is gained from luminesce: which camouflages the squid against the sky, from predators below.

The quorum sensing system of *A. fischeri* comprises of the enzyme LuxI, that catalyses production of the auto-inducer N-(3-oxohexanoyl)-homoserine lactone (3OC6HSL), and the 3OC6HSL inducible transcription factor LuxR, which upon binding 3OC6HSL activates expression from the luxICDABE operon which encodes the components of luciferase (**Figure 1**). Many other quorum sensing systems have been elucidated, which in gram-negative bacteria generally function in the same way, with homologues of LuxI and LuxR, and differing N-homoserine lactone autoinducers.

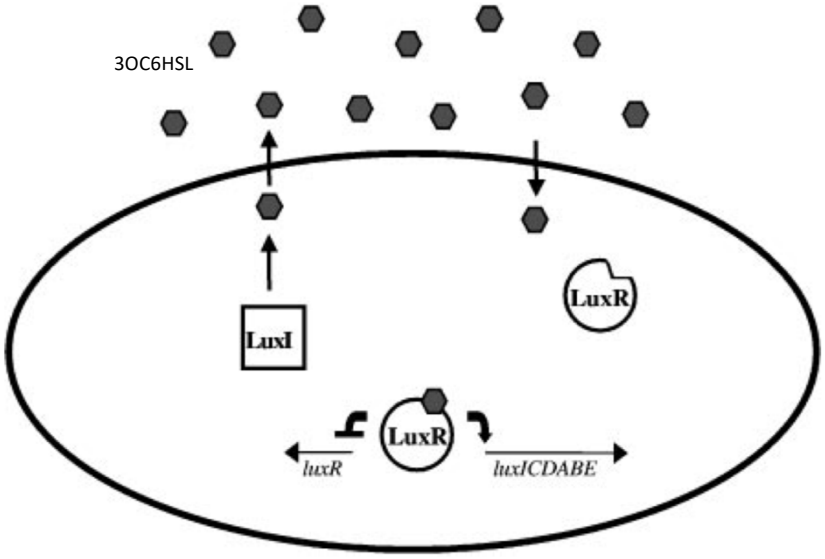

Figure 1. The quorum sensing system of *Aliivibrio fischeri*. Showing The N-acyl homoserine lactone synthase **LuxI** producing N-(3-oxohexanoyl)-homoserine lactone **3OC6HSL** (grey hexagons), which activates the transcription factor **LuxR**, which in turn both activates transcription of the *luxICDABE* operon and represses transcription of *luxR*. Image taken from (12).

Synthetic multicellular networks have already been created utilising QS for inter-cellular signalling (13). For example, Danino and colleagues used QS to synchronise oscillatory bursts of bioluminescence in *Escherichia coli*, demonstrating the ability to use signalling to synchronise cell behaviour (14). Basu and colleagues used QS to produce spatial patterns of fluorescent protein production, using two strains containing different synthetic gene circuits (15).

Feedback Control Systems

In order to have cells autonomously regulating their population density, closed-loop feedback is required (16). Negative feedback regulation is ubiquitous in nature where it is necessary for homoeostasis (17). Organisms constantly need to adapt to changes in internal processes as well as in their environment, and this adaptation is often the result of negative feedback. Negative feedback occurs at all levels of biological complexity from cellular processes to sensory-motor control and even into ecological interactions (17). In yeast cells, changes in turgor pressure at the cell membrane trigger a rise in glycerol synthesis via a signalling cascade. The resulting change in the osmolarity of the cytosol balances the change in the environment and therefore returns turgor pressure to normal (18). Hawk moths constantly use feedback from visual stimuli and alter the shape of their abdomen to control pitch in order to sustain hovering flight while feeding (19). Flapping flight is intrinsically noisy and sensitive to external perturbation from air currents, so without feedback it would be a highly haphazard way to move around (17). At an ecological level, oscillations in predator prey systems resulting from over predation and feedback via prey scarcity are well-documented phenomena (20).

Just as in nature, feedback control is long established in engineering and control theory has developed to provide a formal mathematical framework to study it. In a typical control scheme (**Figure 2**), the output (y) of a controlled system, referred to as the plant, is measured by a sensor. This measured output (ŷ) is compared to an external reference value (r) by a comparator, which then passes the calculated error (e) to a controller. The controller then calculates a correctional signal (û) to be implemented by an actuator as a system input (u) to steer the controlled system back to the desired output level. Control theory is often used in synthetic biology to design stable and robust gene regulatory networks (GRNs) (21). Negative feedback has been used to build oscillators and positive feedback has been used to create bi-stable switches (14, 22, 23). One way to implement feedback regulation in engineered cells is to use an external control system. This approach has been demonstrated in a number of recent studies, in which a control algorithm implemented in a computer was combined with real time *in vivo* measurements to create a closed loop and steer a biological process to a specified level. Milias-Argeitis and colleagues used red and far red light as inputs to trigger PhyB-PIF3 mediated expression of a fluorescent protein in yeast (24). The system output, fluorescence, was measured using flow cytometry and a computer algorithm calculated the control input. Menolascina and colleagues demonstrated a similar system in which time-lapse microscopy was used to measure the fluorescence of yeast cells grown in a microfluidics device (25). A computer algorithm calculated the regulatory input of adding glucose or galactose.

There have also been several examples of synthetic circuits implemented *in vivo to* regulate population density. A single population was engineered by You and collegues (26) to achieve a stable low density over a long period of time, using a QS molecule to activate expression of a toxin protein. Following on from this work, Balagadde and

colleagues created an artificial predator prey system in which two engineered strains interacted via QS molecules to attain stable or oscillatory population dynamics, depending upon media conditions (27). These studies demonstrate the opportunity to utilise QS molecules and toxin-antitoxin proteins to manipulate bacterial cell densities, and to apply this for useful purposes.

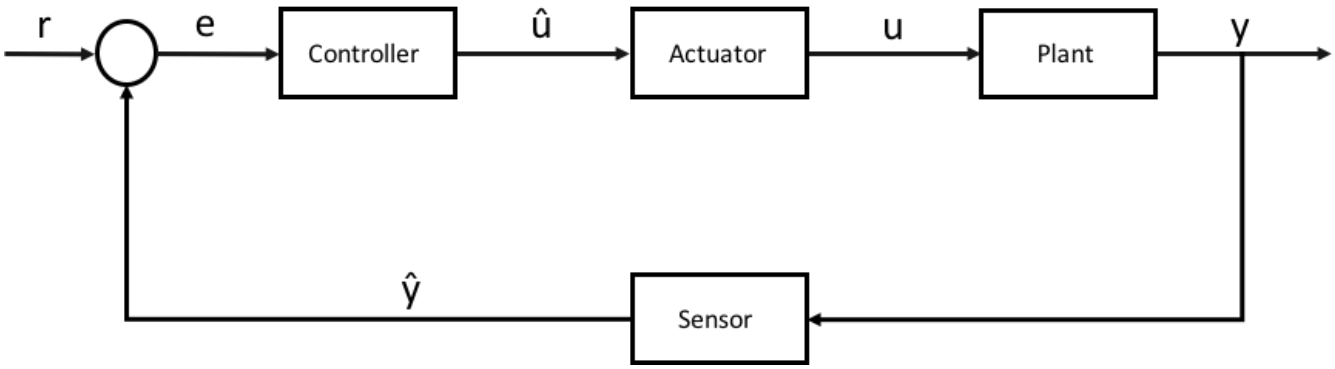

Figure 2. Block diagram of a typical control scheme. Showing the necessary components of a control system and the symbols assigned to inputs and outputs.

Study Aims

Building on the work of You, Balagadde and colleagues (26, 27), the aim of this study was to design a synthetic GRN able to control the population densities of two different bacterial strains. The objective was to have one strain acting as the "master" population, maintaining its own density as well as sensing and controlling the population density of the second "slave" strain, with the additional objective that both population densities are tuneable via an external reference input sensed by the master cells. To achieve these goals, the design uses closed-loop negative feedback implemented within the cells. The interaction between master and slave populations can therefore be viewed in terms of a classical control feedback system with the master population carrying out the sensor, comparator and controller functions and the slave population acting as the plant and containing the actuator. Two populations were modelled as a proof of concept but the general design could be modified for multiple populations cultured together in a synthetic consortium with one population regulating the density of the other populations or alternatively with a cascade of hierarchical control.

RESULTS AND DISCUSSION

Circuit Design

Our circuit design consists of two closed loops (**Figure 3**): the first loop (Labelled **L1** in **Figure 3**) uses QS signalling between master and slave cells and a toxin protein ccdB to create a negative feedback loop regulating slave population density. We used a modification of a circuit designed by Balagadde and colleagues, as its functioning has been confirmed *in vivo* (27). The circuit uses two QS systems: the LuxI/LuxR system from the marine bacterium *A. fischeri* and the LasI/LasR system from *Pseudomonas aeruginosa (28)*. In both systems, the inducer protein (LuxI/LasI) synthesises a small diffusible Homoserine Lactone HSL. The HSL activates its cognate receptor protein (LuxR/LasR) which acts as a transcriptional regulator and increases gene expression from the PluxI promoter (both LuxR and LasR are known to up-regulate expression from PluxI (27), so we used the promoter in both systems). When it accumulates in the cytosol, ccdB protein acts as a poison in *E.coli*, interfering with the DNA gyrase complex (29). In our design, slave cells constitutively express LuxI which synthesises 3OC6HSL, this diffuses into the surrounding media and then into master cells. At sufficient concentration, 3OC6HSL alters LuxR conformation to its active state. LuxR then upregulates expression of LasI; LasI in turn catalyses the synthesis of another HSL: 3OC12HSL, which diffuses into slave cells and activates LasR. Finally, LasR increases expression of ccdB, poisoning a proportion of slave cells, this closes the loop by reducing slave population growth, and therefore 3OC6HSL production. LuxR and LasR are constitutively expressed.

The second loop in the design (Labelled **L2** in **Figure 3**) is required to maintain a stable master population density. Unlike the predator prey system of Balagadde, we needed the master population to regulate its own density independently of slave population density. A similar QS - ccdB negative feedback loop is used to regulate master population density, utilising another QS system, TraI/TraR from *Agrobacterium* species. In this loop however, the strength of the feedback is controllable via an externally inducible promoter. In the presence of arabinose, the transcriptional regulator AraC induces expression from the PBAD promoter (30), therefore TraI expression rate can be controlled via an external arabinose input (araC gene is not included in Figure.2 but is intended to be constitutively present in master cells). Arabinose induces expression of TraI, which synthesises 3OC8HSL, this diffuses into the media and its concentration will depend on master population density. 3OC8HSL will activate TraR, in master cells, which in turn increases expression of ccdB, from the PtraR promoter, reducing master population growth rate. This closes loop 2 by reducing the number of cells producing 3OC8HSL, and similarly reduces the strength of the feedback in loop 1. Therefore, arabinose concentration directly controls master population density and indirectly controls slave population density. Having ccdB expression controlled by 3OC8HSL concentration rather than directly by arabinose concentration allows master cells to dynamically respond to changes in their own population density, whilst simultaneously allowing density to be modulated as desired.

The hypothesis was that this circuit design could perform as a control system in which the master population acts as a controller and slave population density is the system output (**Figure 4**). 3OC6HSL acts as a proxy for slave population density and arabinose acts as the reference signal. The difference between 3OC6HSL and arabinose (via 3OC8HSL) concentration in the media will determine the strength of the 3OC12HSL control signal. ccdB expression in slave cells can be viewed as the actuator of the system.

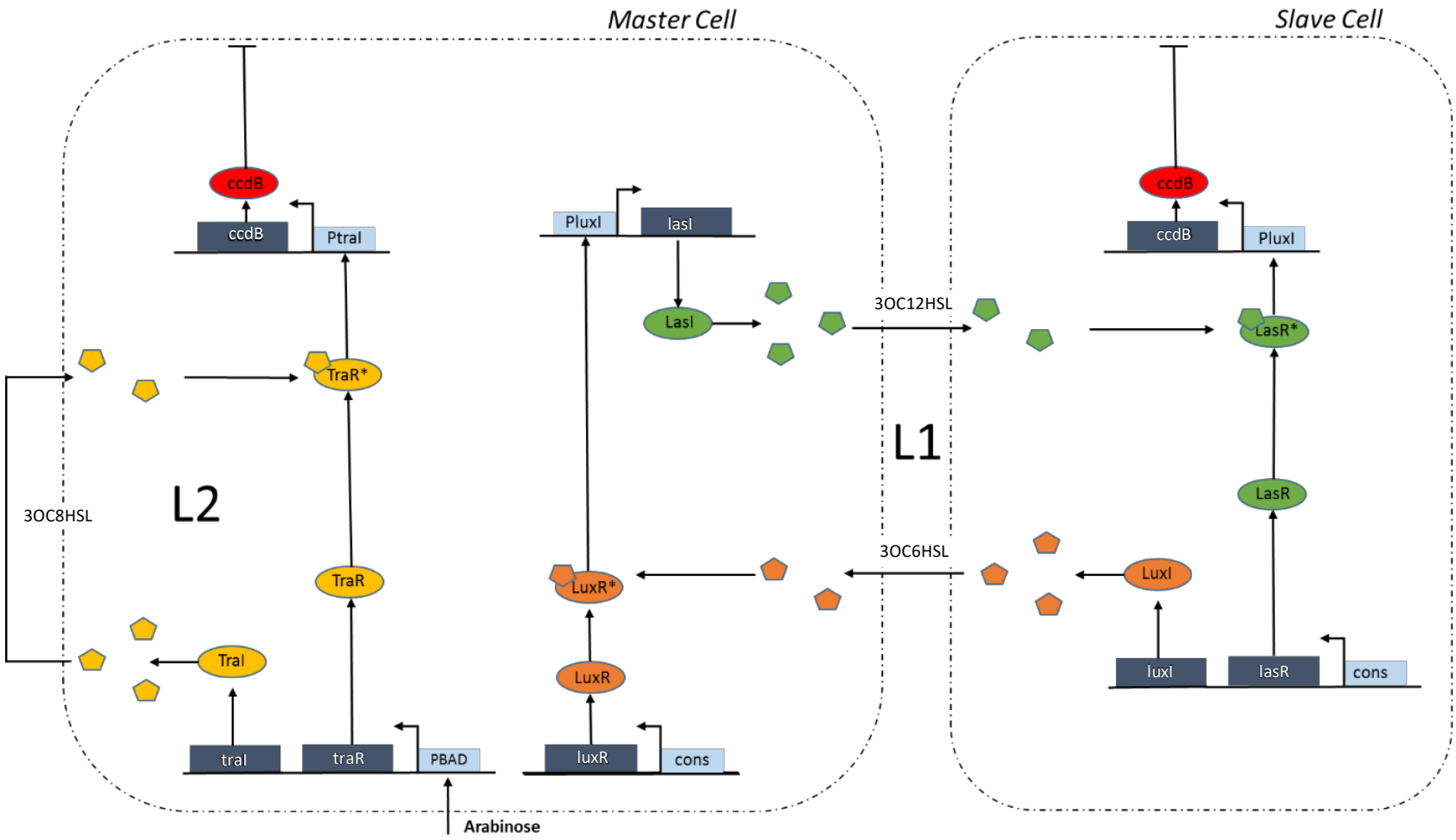

Figure 3. Population density regulating circuit design. The two engineered cell populations are regulated by QS feedback loops that control ccdB toxin expression. The master cell population density can be controlled via arabinose concentration in the growth media. Master cells then regulate slave cell population density via two QS molecules. Dashed lines represent cell walls. Arrows represent activation, synthesis or diffusion; blunt arrows represent killing. Small rectangles represent promoter regions, large rectangles - genes, ellipses - proteins and pentagons - HSL molecules.

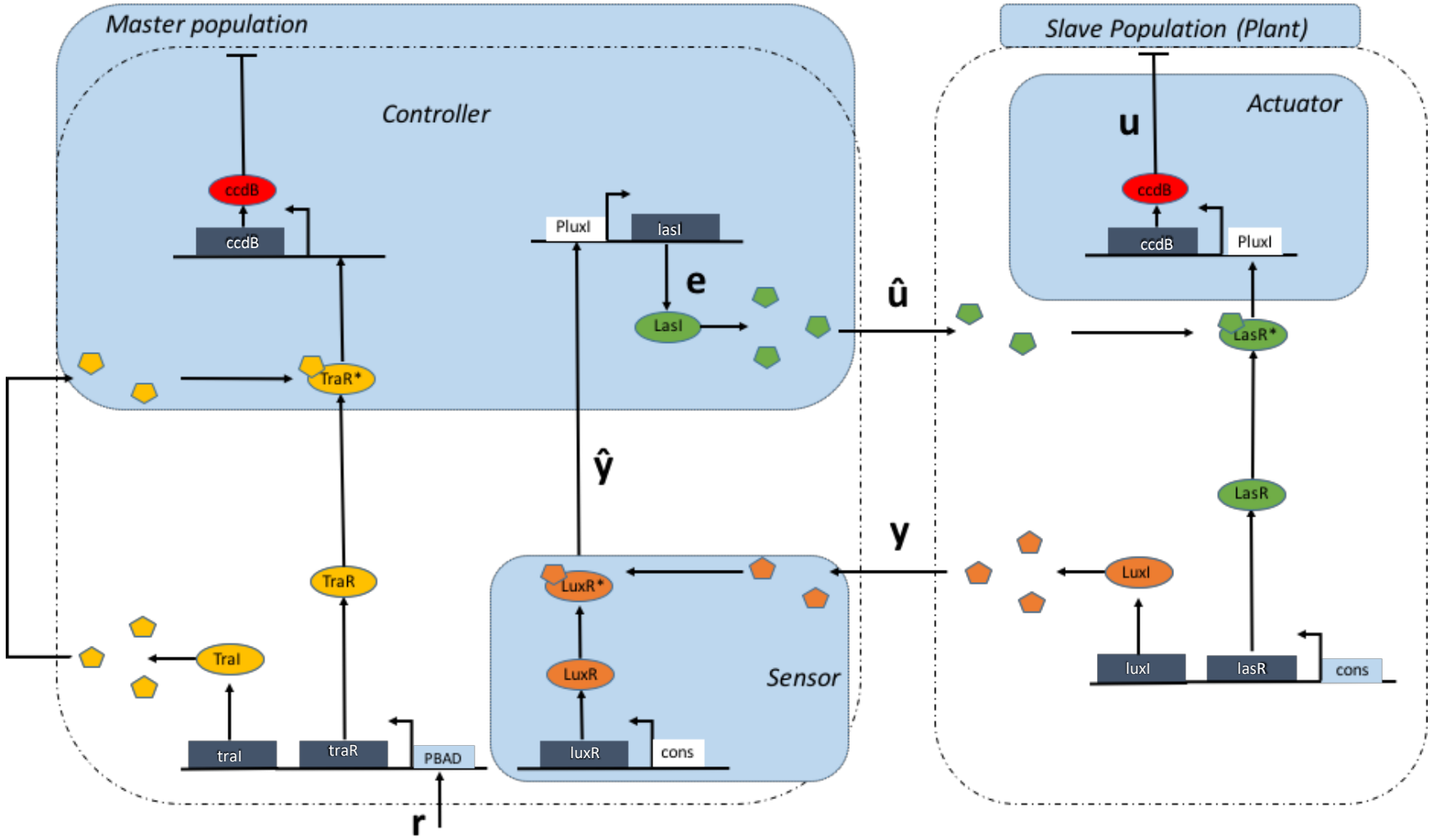

Figure 4. Control system representation of the design. 3OC6HSL acts as a proxy for slave population density and arabinose acts as the reference signal. The difference between 3OC6HSL and arabinose (via 3OC8HSL) concentration in the media will determine the strength of the 3OC12HSL control signal. ccdB expression in slave cells can be viewed as the actuator of the system.

Modelling

Balagadde and colleagues developed a model of all main reactions involved in their predator prey system including transcription, translation, diffusion of HSL and regulatory protein activation; they later simplified this model to equations representing cell population densities and external HSL concentrations only. They concluded that due to large differences in the time scales of processes in the system, this simplification did not greatly affect the behaviour of the model (27). We therefore followed this approach to model our system. The circuit is modelled with a set of five ordinary differential equations (ODEs). These ODEs describe the dynamics of the master population (**Equation 1**) and slave population (**Eq. 2**) and the concentrations of the three HSLs in the media: 3OC6HSL (**Eq. 3**), 3OC8HSL (**Eq. 4**) and 3OC12HSL (**Eq. 5**). Population dynamics are modelled using logistic growth equations including a term that makes cell death proportional to intracellular ccdB concentration. HSL production rates are modelled using Hill function equations, and production rate is assumed to depend directly on the external concentration of any inducer molecule. No intermediate steps including inducer protein transcription and translation, HSL synthesis, HSL diffusion between cells, receptor protein activation and promoter binding are explicitly included. This assumes that all of these processes occur at a faster rate than the modelled processes, and therefore can be removed from the model to reduce complexity. For instance, the equation for 3OC12HSL concentration (**Eq.5**) is written as if 3OC12HSL production is directly activated by 3OC6HSL in the media. The intracellular concentrations of constitutively expressed proteins are treated as being constant, this assumes that these species are in abundance within cells.

Equations. The following equations were used in the model:

$$\frac{dc_1}{dt} = k_{c_1} c_1 \left(1 - \frac{c_1}{c_{max}}\right) - d_{c_1} c_1 \frac{(A_{e8})^{\beta_1}}{K_1 + (A_{e8})^{\beta_1}} - Dc_1 \quad (1)$$

$$\frac{dc_2}{dt} = k_{c_2} c_2 \left(1 - \frac{c_2}{c_{max}}\right) - d_{c_2} c_2 \frac{(A_{e12})^{\beta_2}}{K_2 + (A_{e12})^{\beta_2}} - Dc_2 \quad (2)$$

$$\frac{dA_{e6}}{dt} = k_{A_6} c_2 - \left(d_{A_{e6}} + D\right) A_{e6} \quad (3)$$

$$\frac{dA_{e8}}{dt} = k_{A_8} c_1 \frac{(I_e)^{\beta_3}}{K_3 + (I_e)^{\beta_3}} - \left(d_{A_{e8}} + D\right) A_{e8} \quad (4)$$

$$\frac{dA_{e12}}{dt} = k_{A_{12}} c_1 \frac{(A_{e6})^{\beta_4}}{K_4 + (A_{e6})^{\beta_4}} - \left(d_{A_{e12}} + D\right) A_{e12} \quad (5)$$

Where:

Subscript 1 refers to the master cells, and 2 refers to the slave cells, subscripts 6, 8 and 12 refer to 3OC6HSL, 3OC8HSL and 3OC12HSL respectively;

c_1 is the master population density (cells nl^{-1});

c_2 is the slave population density (cells nl^{-1});

k_{c_i} is the growth rate of master/slave cells (hour^{-1});

c_{max} is the carrying capacity of the system;

d_{c_i} is the death rate of master/slave cells (nM^{-1} ccdB hour^{-1});

A_{ei} is the external HSL (3OCiHSL) concentration (nM);

β_i is a cooperativity (Hill) coefficient;

K_i is the external concentration of the relevant signal molecule needed to half maximally activate its cognate promoter (nM);

D is the dilution rate (hour^{-1});

k_{A_i} is the production rate of HSL (3OCiHSL) (nM hour^{-1});

d_{A_i} is the external degradation rate of HSL (3OCiHSL) (hour^{-1}); and

I_e is the external arabinose concentration (nM).

Parameter values. **Table 1** contains a full list of parameter values used for modelling. Most initial estimates were taken directly from Balagadde (27) but averaged or rounded values from Balagadde were also used. Other estimates were made using consensus values from the literature.

Table 1. Parameter values

Parameter	Description	Unit	Original value	Final value
k_{c_1}	Master population growth rate	hr^{-1}	0.5	/
k_{c_2}	Slave population growth rate	hr^{-1}	0.25	/
d_{c_1}	Master cell death rate	nM^{-1} ccdB hr^{-1}	0.5	/
d_{c_2}	Slave cell death rate	nM^{-1} ccdB hr^{-1}	0.25	/
c_{max}	System carrying capacity	cells nL^{-1}	10^5	/
D	Dilution rate	hr^{-1}	0.3	/
β_1	Cooperatively of 3OC8HSL activation	NA	2	/
β_2	Cooperatively of 3OC12HSL activation	NA	2	/
β_3	Cooperatively of Arabinose activation	NA	2	/
β_4	Cooperatively of 3OC6HSL activation	NA	2	/
K_1	Concentration of 3OC8HSL necessary to half-maximally active PtraI promoter	nM	10	/
K_2	Concentration of 3OC12HSL necessary to half-maximally active PluxI promoter	nM	10	/
K_3	Concentration of Arabinose necessary to half-maximally active PBAD promoter	nM	10^5	10^{13}
K_4	Concentration of 3OC6HSL necessary to half-maximally active PluxI promoter	nM	10	/
k_{A_6}	Production rate of 3OC6HSL	nM hr^{-1}	0.1	0.001
k_{A_8}	Production rate of 3OC8HSL	nM hr^{-1}	0.1	/
$k_{A_{12}}$	Production rate of 3OC12HSL	nM hr^{-1}	0.1	0.001
$d_{A_{e6}}$	External degradation rate of 3OC6HSL	hr^1	0.05	0.0005
$d_{A_{e8}}$	External degradation rate of 3OC8HSL	hr^1	0.05	/
$d_{A_{e12}}$	External degradation rate of 3OC12HSL	hr^1	0.05	0.0005

Circuit Functioning

Using the original parameter estimates, with HSL production and degradation rates and activation coefficients set equal (**Table.1**), slave population density crashed at arabinose concentrations in the μM – mM range (**Figure 5**). In order to allow slave population density to rise, 3OC6HSL and 3OC12HSL production rates (k_{A_6},$k_{A_{12}}$) were reduced 10 or 100 fold with HSL degradation rates ($d_{A_{e6}}$,$d_{A_{e12}}$) set at half production rate. Reducing HSL production rates 10 fold allowed the slave population to reach higher steady state (s.s.) density (≈ 40,000 cells nL^{-1}) (**Figure 5)**. Reducing production rates 100 fold had the extra advantage of reducing the time taken to reach equilibrium, so these rates were used for all further analyses. These reductions in production rates could be achieved in reality by reducing the relevant plasmid copy numbers or by selecting very weak promoters. A bifurcation plot (using the adjusted parameter set) showed that while slave population now reached a high, stable density, density could not be controlled by altering arabinose concentration within the realistic concentration range (**Figure 6**). In order to control slave population density, the sensitivity of the PBAD promoter to arabinose had to be reduced. As the arabinose activation coefficient (K_3) was increased from 10^5, the control of slave population density using the desired arabinose concentrations becomes possible (**Figure 7**). For the following analysis $K_3 = 10^{13}$ was used. Some decrease in activation sensitivity could be achieved using a mutated transcription factor or degenerated promoter to reduce affinity. However, it is uncertain if an activation coefficient as high as this could be achieved in reality. With the adjusted parameter set, master and slave population densities respond to arabinose concentration achieving the study design aims (**Figures 8, 9 and 10**). Populations settle at stable densities, and the densities reached can be altered via arabinose concentration. S.s. Slave population density can be tuned from less than 1000 (**Figure 8**) to more than 30,000 cells nL^{-1} (**Figure 10**) using arabinose concentrations of 0.1 to 1 mM.

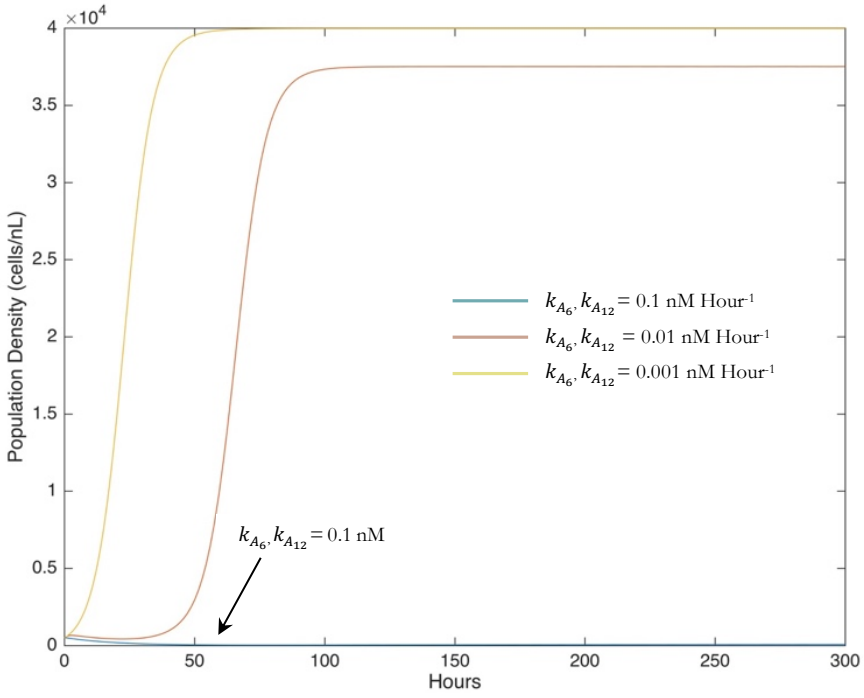

Figure 5. Slave population dynamics at different HSL production rates. Production rates k_{A_6} and $k_{A_{12}}$ at 0.1, 0.01 and 0.001 nM Hour^{-1} ($d_{A_{e6}}$ and $d_{A_{e12}}$ were set to equal half production rate). Only the lower production rates allow the slave population to reach high densities.

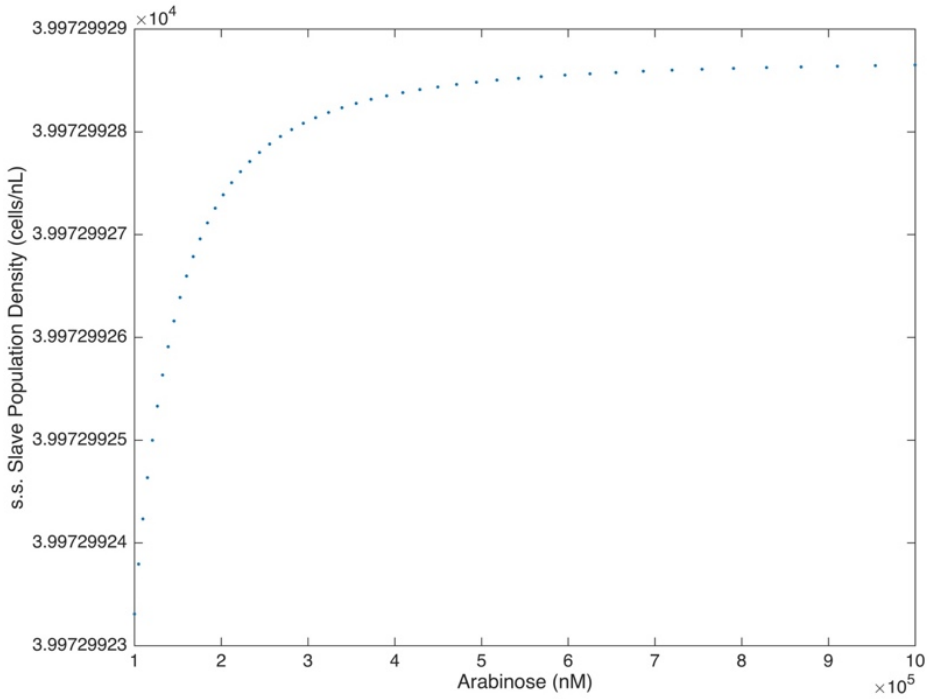

Figure 6. Steady state slave population density against arabinose concentration with original parameters. Bifurcation plot, with k_{A_6} and $k_{A_{12}}$ = 0.001 nM Hour^{-1}), showing that s.s. population density cannot be controlled sufficiently via arabinose concentrations in this range.

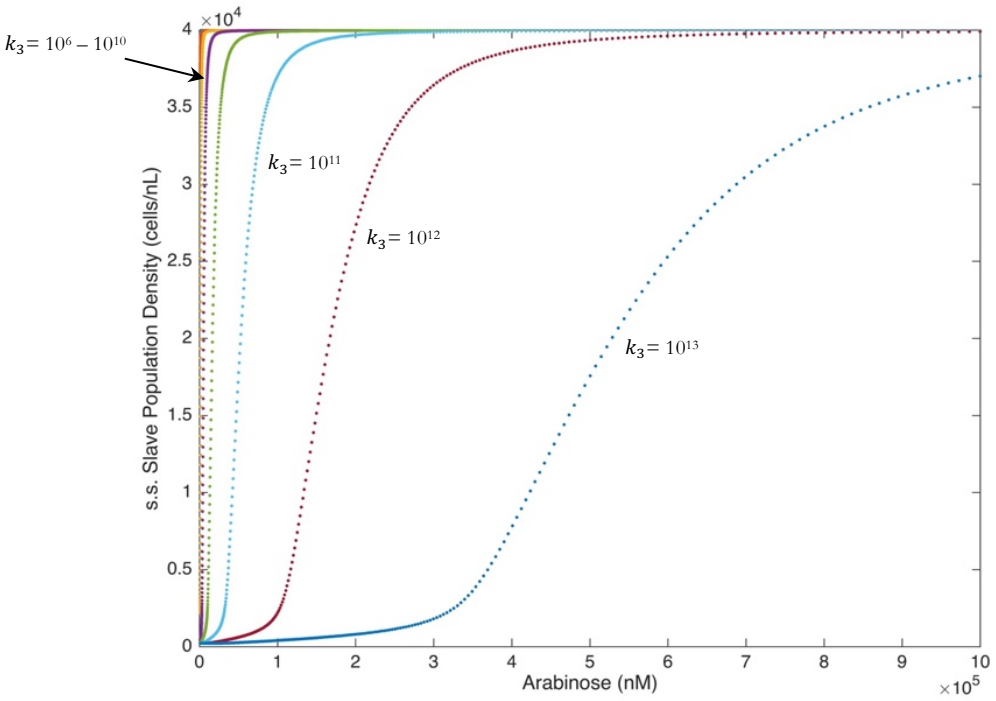

Figure 7. Steady state slave population density against arabinose concentration: altered activation coefficient. Bifurcation plot, with activation coefficient K_3 = 10^6-10^{13}, showing that K_3 = 10^{13} allows control of slave population densities over an order of magnitude with the selected arabinose concentration range.

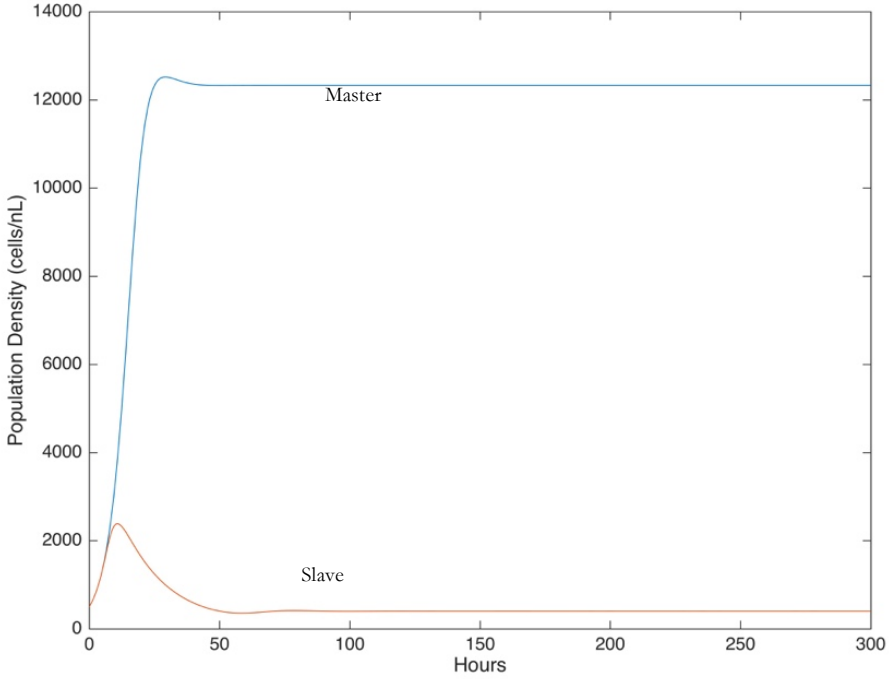

Figure 8. Model population dynamics with arabinose = 0.1 mM. At this arabinose concentration, and with the functional parameter set, Master (Blue) and Slave (Red) populations reach stable densities of approximately 12,000 and 500 cells nL^{-1} respectively.

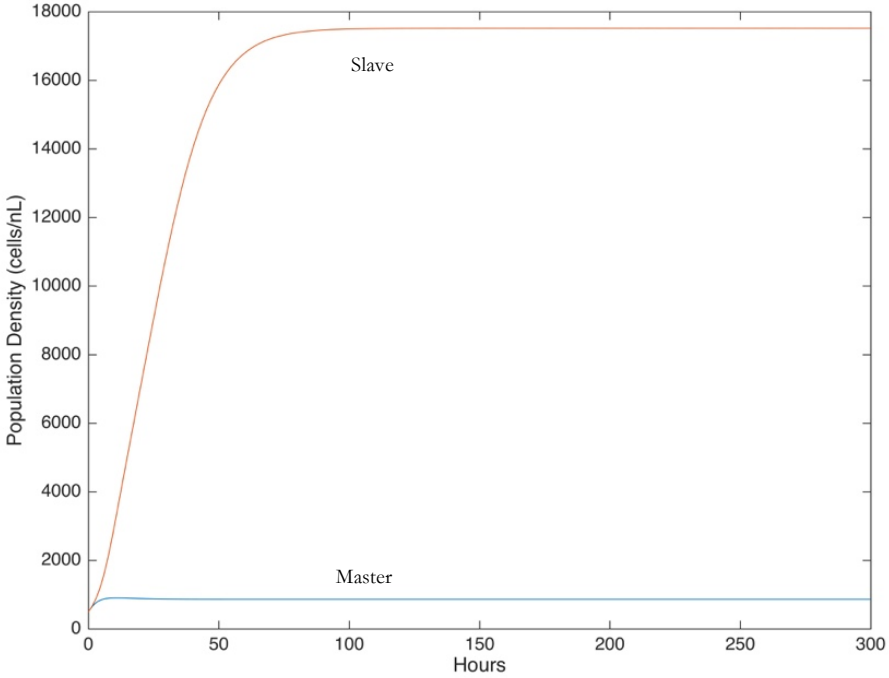

Figure 9. Model population dynamics: Arabinose = 0.5 mM. At this arabinose concentration, and with the functional parameter set, Master (Blue) and Slave (Red) populations reach stable densities of approximately 750 and 17,500 cells nL^{-1} respectively.

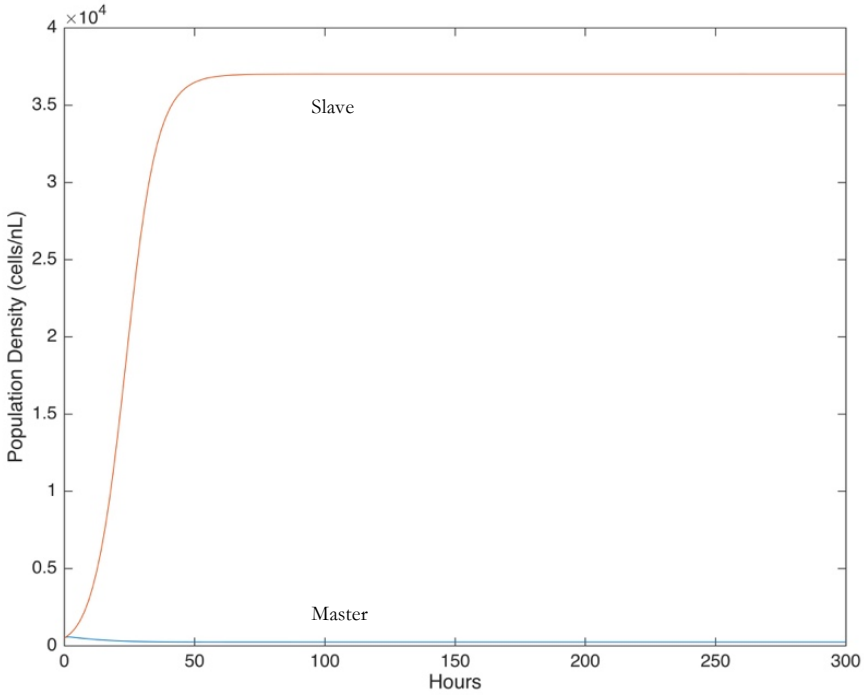

Figure 10. Model population dynamics: Arabinose = 1 mM. At this arabinose concentration, and with the functional parameter set, Master (Blue) and Slave (Red) populations reach stable densities of approximately 500 and 37,000 cells nL^{-1} respectively.

Sensitivity Analysis

To ascertain which parameters are most sensitive to change, bifurcation analysis was carried out for each parameter in turn. The effect on s.s. slave population density of altering each parameter for a range of ±10% from the original value was determined with arabinose concentration = 0.6 mM which corresponds to a s.s. slave density of 25,000 cells nL^{-1} (**Figures 11-22**). This provided an indication of the robustness of the system to noise and identified which parameters are potential targets for manipulation to improve system performance. For parameters where values were all equal — cooperativity coefficients (β_1, β_2, β_3, and β_4), HSL production rates ($k_{A_6}, k_{A_{12}}$) and HSL degradation rates ($d_{A_{e6}}, d_{A_{e12}}$) — all parameters were perturbed simultaneously. The following results were observed:

1. As can be expected intuitively from the design, the following parameters are positively correlated with s.s. slave population density: master cell death rate (d_{c_1}) (**Figure 12**), slave population growth rate (k_{c_2}) (**Figure 13**), cooperativity/Hill coefficients (β_1, β_2, β_3, and β_4) (**Figure 15**), 3OCHSL6 and 3OCHSL12 degradation rate ($d_{A_{e6}}, d_{A_{e12}}$) (**Figure 19**), 3OCHSL8 production rate (k_{A_8}) (**Figure 20**) and population carrying capacity (c_{max}) (**Figure 22**). While the following are negatively correlated: master population growth rate (k_{c_1}) (**Figure 11**), slave cell death rate (d_{c_2}) (**Figure 14**), activation coefficients (K_1, K_2, K_4) (**Figure 16**), 3OCHSL6 and 3OCHSL12 production rate (k_{A_6},$k_{A_{12}}$) (**Figure 18**) and 3OCHSL8 degradation rate ($d_{A_{e8}}$) (**Figure 21**).

2. Master population growth rate (k_{c_1}) is the parameter most sensitive to change (**Figure 11**), decreasing by 10% has limited effect on s.s. slave population density, but a 10% increase results in an approximately 10-fold reduction in s.s. slave density. This suggests that when selecting a master strain, erring on the side of low growth rate would be wise. Other population change parameters (k_{c_2},d_{c_1}, d_{c_2}) only exert a comparatively moderate effect on s.s. slave population density.

3. s.s. Slave population density is highest for a dilution rate of 0.31, and increasing or decreasing *D* lowers s.s. population density (**Figure 17**).

4. All other parameters have a small or moderate effect on s.s. slave density and so are relatively robust to estimation error or natural variation.

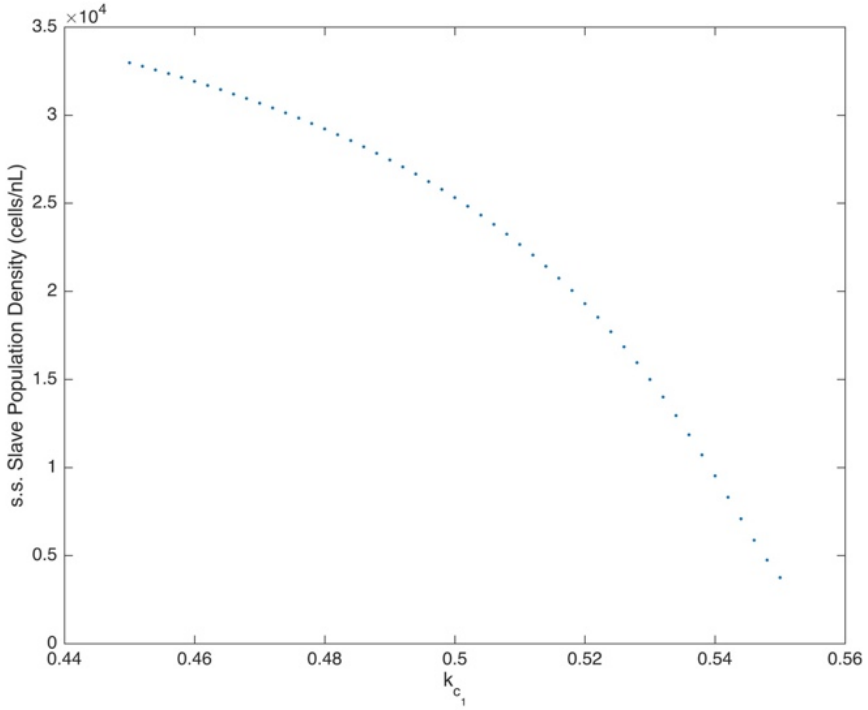

Figure 11. Parameter sensitivity analysis: Master population growth rate. Bifurcation plot showing the effect on steady state slave population density of altering Master population growth rate (k_{c_1}), for a range ±10% from the original value.

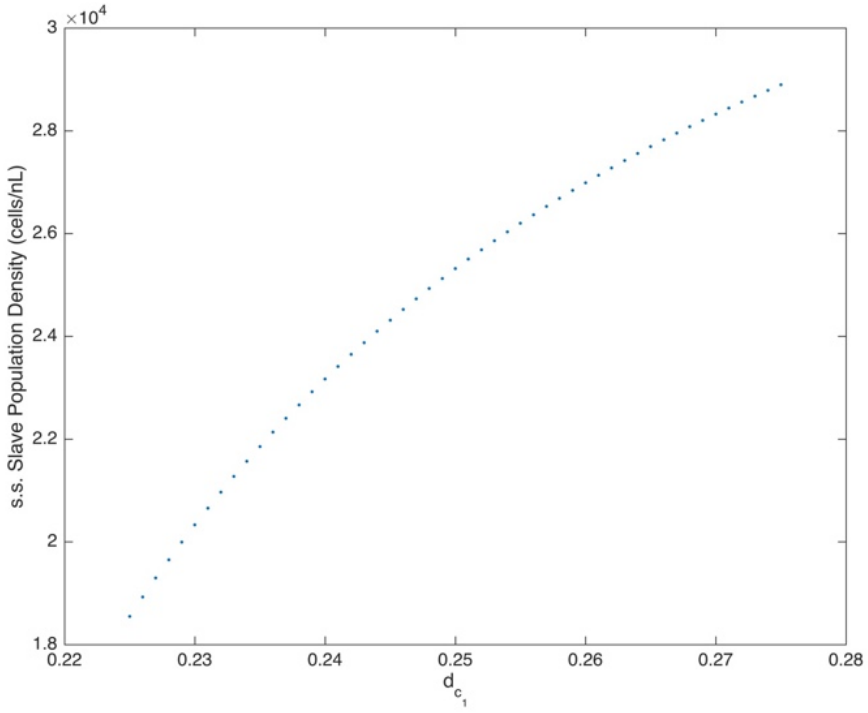

Figure 12. Parameter sensitivity analysis: Master cell death rate. Bifurcation plot showing the effect on steady state slave population density of altering Master cell death rate (d_{c_1}), for a range ±10% from the original value.

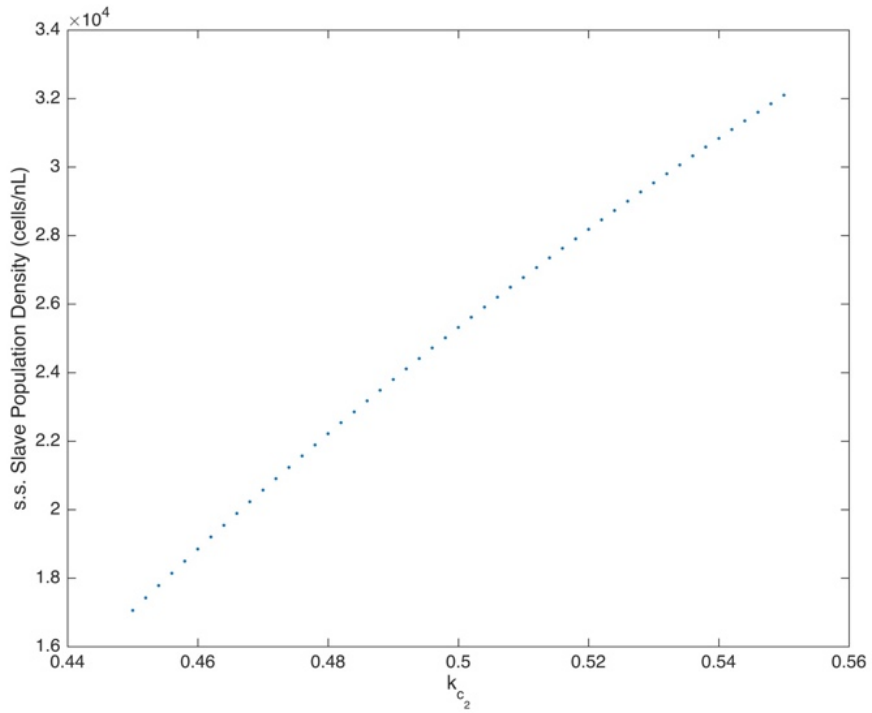

Figure 13. Parameter sensitivity analysis: Slave population growth rate. Bifurcation plot showing the effect on steady state slave population density of altering Slave population growth rate (k_{c_2}), for a range ±10% from the original value.

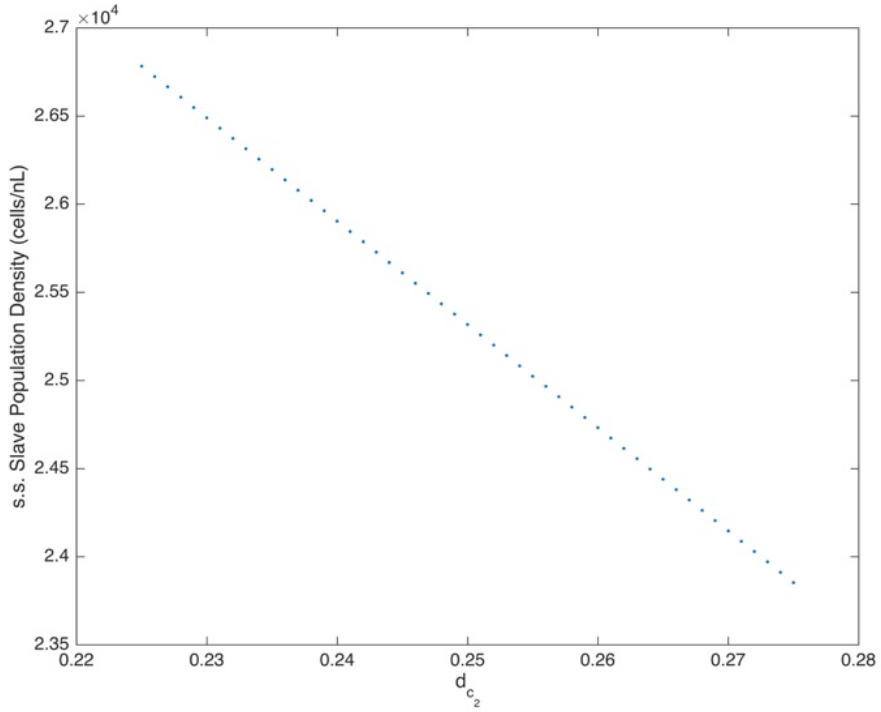

Figure 14. Parameter sensitivity analysis: Slave cell death rate. Bifurcation plot showing the effect on steady state slave population density of altering Slave cell death rate (d_{c_2}), for a range ±10% from the original value.

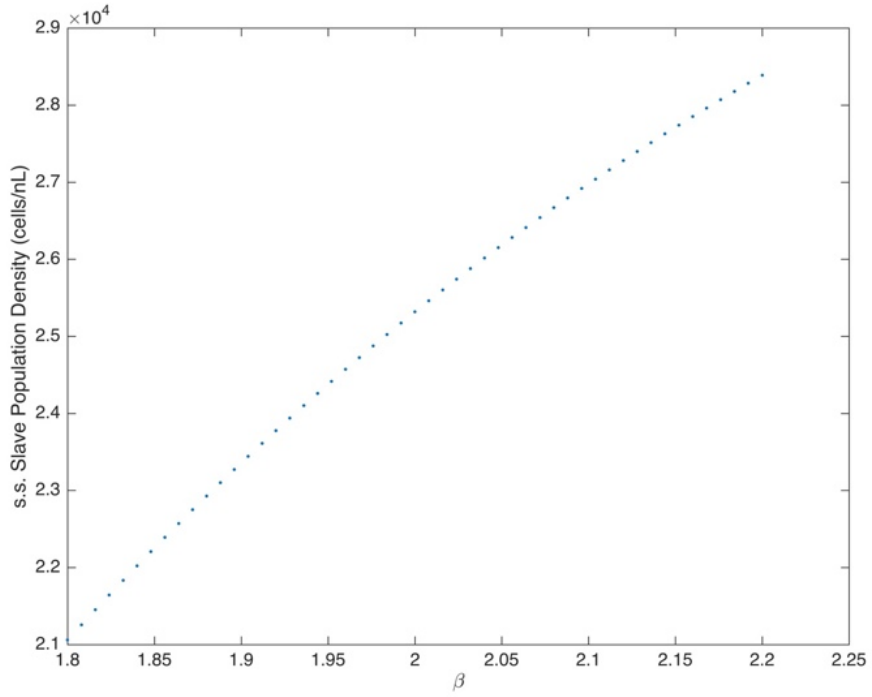

Figure 15. Parameter sensitivity analysis: Cooperatively coefficients. Bifurcation plot showing the effect on steady state slave population density of altering Cooperatively coefficients *(β_1, β_2, β_3,* and *β_4)*, for a range ±10% from the original value.

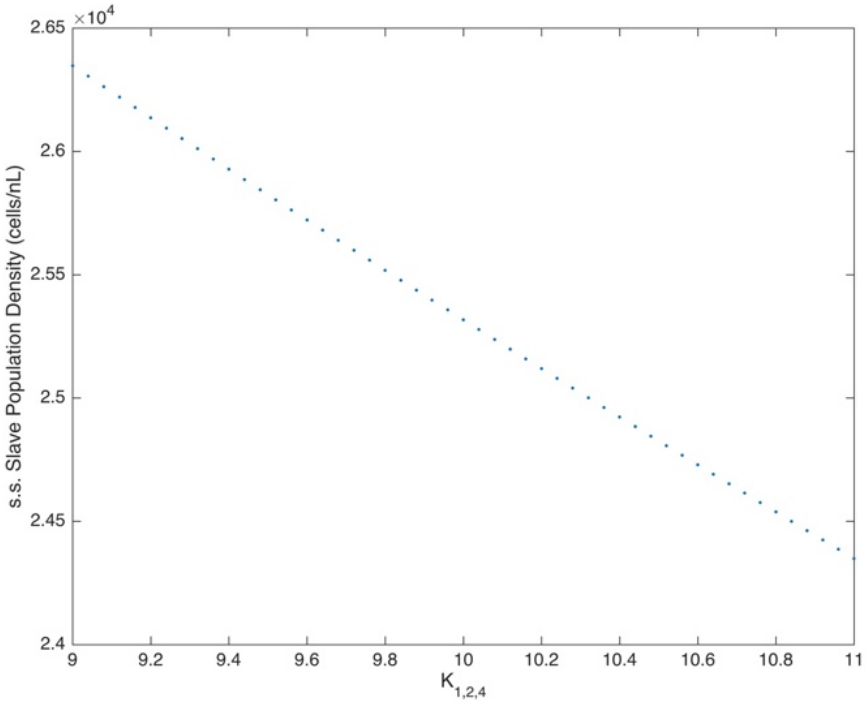

Figure 16. Parameter sensitivity analysis: Activation coefficients. Bifurcation plot showing the effect on steady state slave population density of altering Activation coefficients (*K_1, K_2, K_4*), for a range ±10% from the original value.

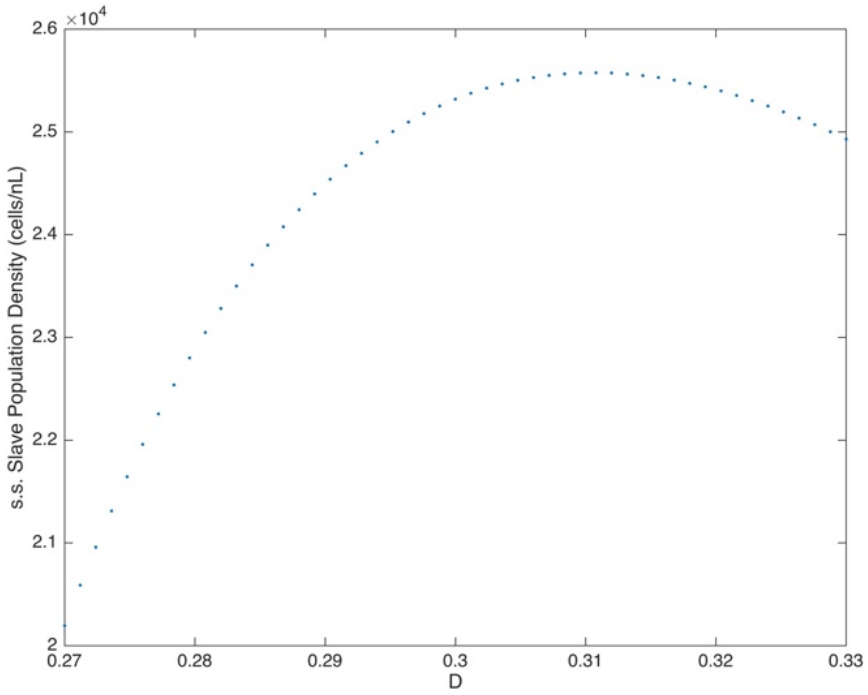

Figure 17. Parameter sensitivity analysis: Dilution rate. Bifurcation plot showing the effect on steady state slave population density of altering Dilution rate (*D*), for a range ±10% from the original value.

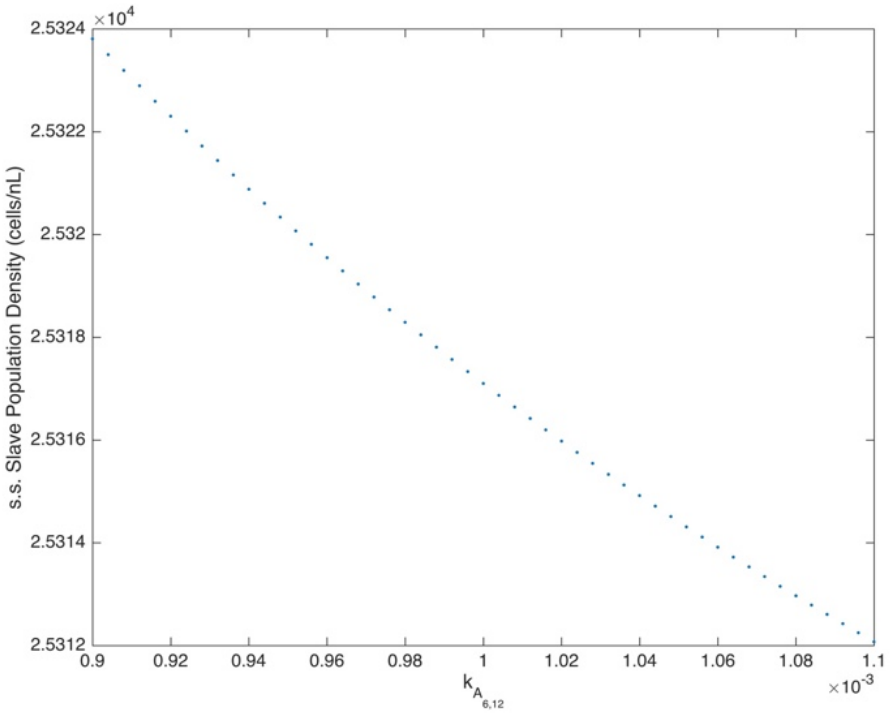

Figure 18. Parameter sensitivity analysis: 3OCHSL6 and 3OCHSL12 production rate. Bifurcation plot showing the effect on steady state slave population density of altering 3OCHSL6 and 3OCHSL12 production rate (k_{A_6}, $k_{A_{12}}$), for a range ±10% from the original value.

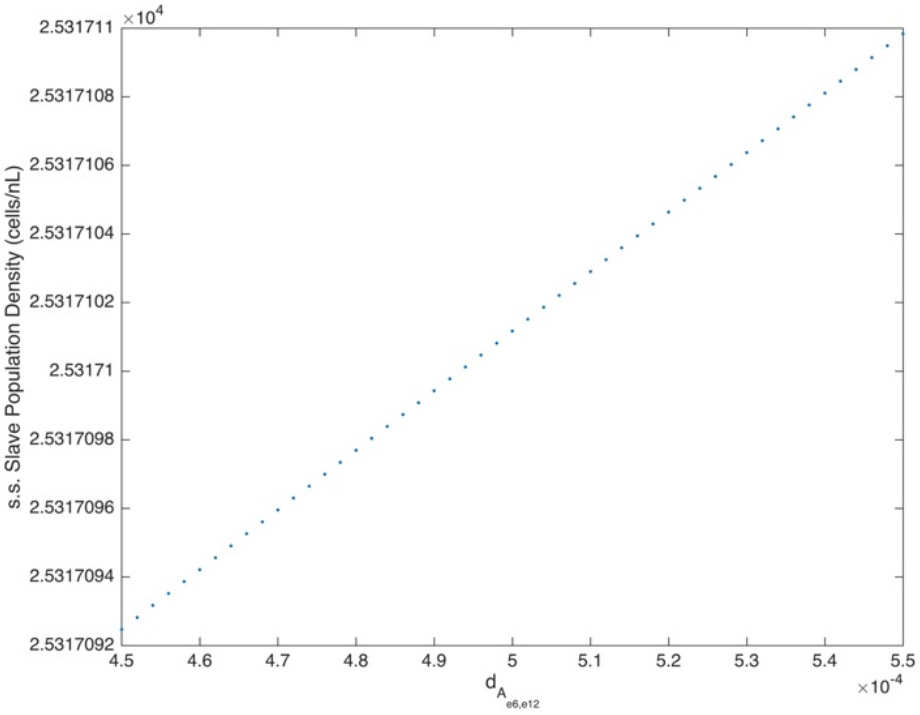

Figure 19. Parameter sensitivity analysis: 3OCHSL6 and 3OCHSL12 degradation rate. Bifurcation plot showing the effect on steady state slave population density of altering 3OCHSL6 and 3OCHSL12 degradation rate ($d_{A_{e6}}$, $d_{A_{e12}}$), for a range ±10% from the original value.

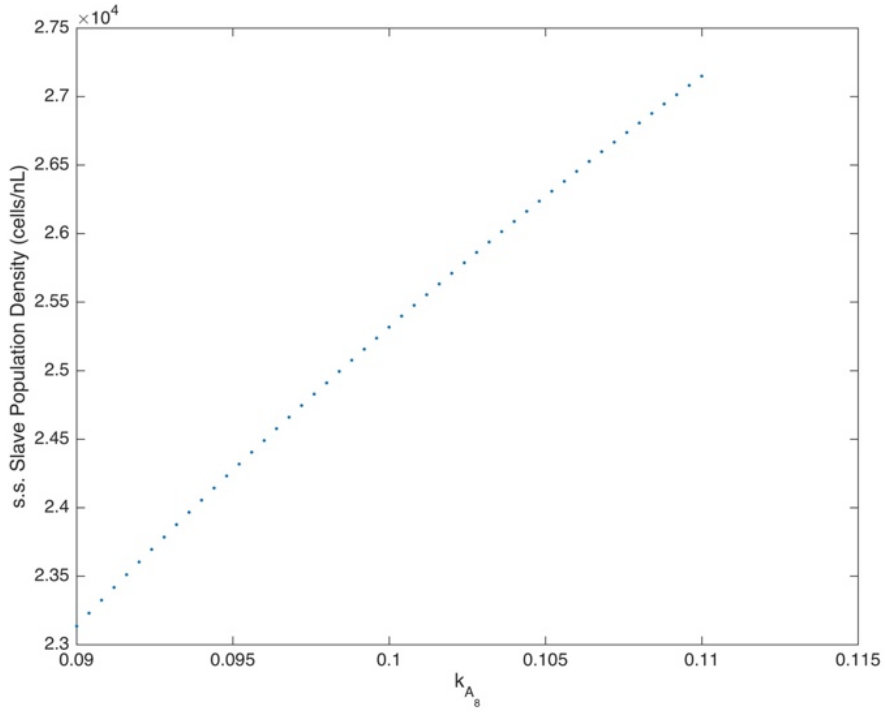

Figure 20. Parameter sensitivity analysis: 3OCHSL8 production rate. Bifurcation plot showing the effect on steady state slave population density of altering 3OCHSL8 production rate (k_{A_8}), for a range ±10% from the original value.

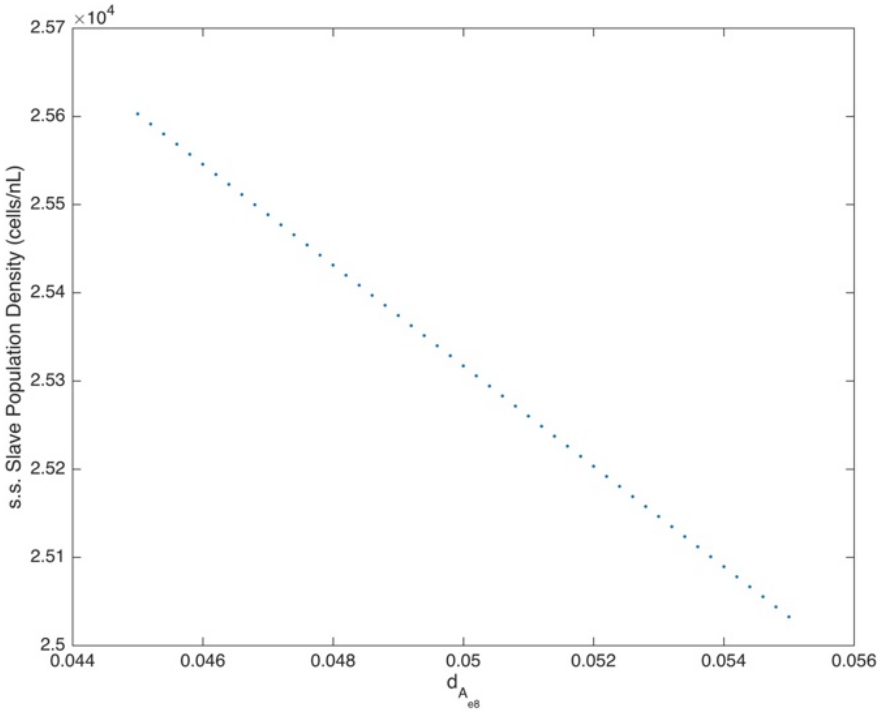

Figure 21. Parameter sensitivity analysis: 3OCHSL8 degradation rate. Bifurcation plot showing the effect on steady state slave population density of altering 3OCHSL8 degradation rate ($d_{A_{e8}}$), for a range ±10% from the original value.

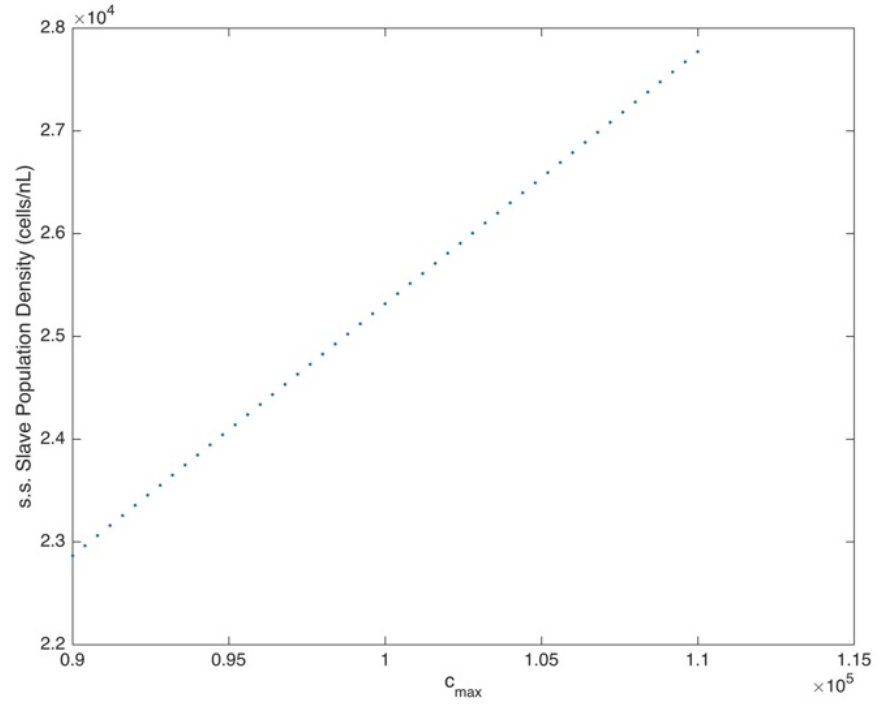

Figure 22. Parameter sensitivity analysis: Population carrying capacity. Bifurcation plot showing the effect on steady state slave population density of altering Population carrying capacity (c_{max}), for a range ±10% from the original value.

CONCLUSIONS

The simulation demonstrates that this circuit design for regulating bacterial populations can function at least theoretically. Modelling results show that our simple design can perform the specified aims of maintaining stable, but different, densities of two populations and allowing s.s. densities to be tuneable through an external input. Parameter analyses provide some insight into functional parameter ranges and potential targets to modify performance. To put the design into practice, the genetic components of the circuit could be assembled on three plasmids: a TraI, TraR and ccdB expression plasmid for self-regulation, contained within the master cells; a LuxR and LasI slave regulating plasmid, contained within master cells; and a LuxI, LasR and ccdB plasmid, within slave cells. This conformation would provide modularity allowing only the components required for a given task to be used. For example, a single population system would only require the self-regulation plasmid. This design could also be extended to regulate multiple populations by introducing variants of the slave plasmid into multiple slave strains, a consortium with many strains existing at various densities could be created. Similar circuits could be designed for yeast and mammalian cells, by using pheromones and growth factors or secondary messengers, respectively, for inter-cellular communication (7).

In theory, circuits like this could allow us to create multi-population systems that function autonomously and can adapt to variation in conditions. There are many potential applications of a functioning circuit of this type. As a concrete example, Goyal and colleagues used a consortium of four yeast strains to produce ethanol from cellulose. To produce optimal yield, strains had to be cultured individually before being mixed at specific ratios (31). For a bioprocess like this to be industrially feasible, a means of keeping the ratio stable over a long period is clearly necessary. There have been several reports of cell signalling systems being co-opted to facilitate microbial coexistence. While our system relies on interactions that decrease survival, other researchers have modelled and built systems that use mutual rescue to force obligate mutualism and co-survival of strains (32, 33). It would be interesting to compare the properties of each approach. Yet another way of using multiple strains to carry out a process is to create spatial or temporal separation. Each solution will have its own pros and cons and may be best suited to certain applications.

Our design is very simple and suffers from some limitations. For instance, the exact s.s. population densities for a given arabinose input cannot be known beforehand and would need to be determined experimentally. A more tightly controlled system could be created by adding a circuit designed to act as a direct comparator, to compare the reference signal with active TraR or LuxR concentration, but this would also increase the complexity of the circuit, which entails its own problems. In our design, the ratio between master and slave populations and absolute population densities cannot be controlled independently. One way to alter this would be to introduce another external induction molecule to regulate the strength of master feedback to slave cells, but this would perhaps negate the aim of having interacting populations. Sensitivity analysis demonstrated that deviations from estimated parameter values, especially master population growth rate, greatly affect the outcome and therefore care would be needed to keep growth conditions constant. Other parameters were not as sensitive to slight alteration however, showing that

overall the design is robust. One limitation of all multi-cell synthetic circuits is the lack of available signalling molecules and this is a problem that will need to be addressed to create complex networks.

Towards the long-term goal of creating synthetic microbial consortia with stable populations, our next objective is to assemble and test this network design *in vivo*. Our design intentionally uses circuit parts that have already been assembled and demonstrated working. An initial goal would be to show stable populations in a microfluidics device or micro-chemostat, before scaling up to larger cultures. Research into synthetic consortia is an exciting field, and has innumerable potential applications. An interesting parallel is occurring in swarm robotics, where instead of creating single complex robots, multiple interacting independent agents, each following simple algorithm, have been shown to produce complex emergent behaviours (34). We may therefore expect to see many promising developments in multi-population synthetic networks.

ABBREVIATIONS

QS, Quorum Sensing; GRN, Gene Regulatory Network; HSL, Homoserine Lactone; S.S., Steady State; ODE, Ordinary Differential Equation.

REFERENCES

1. Andrianantoandro E, Basu S, Karig DK, Weiss R. Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems Biology. 2006;2:14.

2. Paddon CJ, Keasling JD. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nature Reviews Microbiology. 2014;12(5):355-67.

3. Wang BJ, Kitney RI, Joly N, Buck M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nature Communications. 2011;2:9.

4. Anderson JC, Clarke EJ, Arkin AP, Voigt CA. Environmentally controlled invasion of cancer cells by engineered bacteria. Journal of Molecular Biology. 2006;355(4):619-27.

5. Pardee K, Green AA, Ferrante T, Cameron DE, DaleyKeyser A, Yin P, et al. Paper-Based Synthetic Gene Networks. Cell. 2014;159(4):940-54.

6. Hennig S, Rodel G, Ostermann K. Artificial cell-cell communication as an emerging tool in synthetic biology applications. Journal of Biological Engineering. 2015;9:12.

7. Bulter T, Lee SG, Woirl WWC, Fung E, Connor MR, Liao JC. Design of artificial cell-cell communication using gene and metabolic networks. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(8):2299-304.

8. Purnick PEM, Weiss R. The second wave of synthetic biology: from modules to systems. Nature Reviews Molecular Cell Biology. 2009;10(6):410-22.

9. Amos M. Population-based microbial computing: a third wave of synthetic biology? International Journal of General Systems. 2014;43(7):770-82.

10. Bagh S, Mandal M, Ang J, McMillen DR. An Active Intracellular Device to Prevent Lethal Disease Outcomes in Virus-infected Bacterial Cells. Biotechnology and Bioengineering. 2011;108(3):645-54.

11. Menendez DB, Senthivel VR, Isalan M. Sender-receiver systems and applying information theory for quantitative synthetic biology. Current Opinion in Biotechnology. 2015;31:101-7.

12. Miller MB, Bassler BL. Quorum sensing in bacteria. Annual Review of Microbiology. 2001;55:165-99.

13. Weiss LE, Badalamenti JP, Weaver LJ, Tascone AR, Weiss PS, Richard TL, et al. Engineering motility as a phenotypic response to LuxI/R-dependent quorum sensing in Escherichia coli. Biotechnology and Bioengineering. 2008;100(6):1251-5.

14. Danino T, Mondragon-Palomino O, Tsimring L, Hasty J. A synchronized quorum of genetic clocks. Nature. 2010;463(7279):326-30.

15. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. A synthetic multicellular system for programmed pattern formation. Nature. 2005;434(7037):1130-4.

16. Chen SS, Harrigan P, Heineike B, Stewart-Ornstein J, El-Samad H. Building robust functionality in synthetic circuits using engineered feedback regulation. Current Opinion in Biotechnology. 2013;24(4):790-6.

17. Cowan NJ, Ankarali MM, Dyhr JP, Madhav MS, Roth E, Sefati S, et al. Feedback Control as a Framework for Understanding Tradeoffs in Biology. Integrative and Comparative Biology. 2014;54(2):223-37.

18. Varela JCS, Praekelt UM, Meacock PA, Planta RJ, Mager WH. THE SACCHAROMYCES-CEREVISIAE HSP12 GENE IS ACTIVATED BY THE HIGH-OSMOLARITY GLYCEROL PATHWAY AND

NEGATIVELY REGULATED BY PROTEIN-KINASE-A. Molecular and Cellular Biology. 1995;15(11):6232-45.

19. Dyhr JP, Morgansen KA, Daniel TL, Cowan NJ. Flexible strategies for flight control: an active role for the abdomen. Journal of Experimental Biology. 2013;216(9):1523-36.

20. Abrams PA. The evolution of predator-prey interactions: Theory and evidence. Annual Review of Ecology and Systematics. 2000;31:79-105.

21. Cury JER, Baldissera FL. Systems Biology, Synthetic Biology and Control Theory: A promising golden braid. Annual Reviews in Control. 2013;37(1):57-67.

22. Muller K, Engesser R, Metzger S, Schulz S, Kampf MM, Busacker M, et al. A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. Nucleic Acids Research. 2013;41(7):11.

23. Garcia-Ojalvo J, Elowitz MB, Strogatz SH. Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(30):10955-60.

24. Milias-Argeitis A, Summers S, Stewart-Ornstein J, Zuleta I, Pincus D, El-Samad H, et al. In silico feedback for in vivo regulation of a gene expression circuit. Nature Biotechnology. 2011;29(12):1114-6.

25. Menolascina F, Fiore G, Orabona E, De Stefano L, Ferry M, Hasty J, et al. In-Vivo Real-Time Control of Protein Expression from Endogenous and Synthetic Gene Networks. Plos Computational Biology. 2014;10(5):14.

26. You LC, Cox RS, Weiss R, Arnold FH. Programmed population control by cell-cell communication and regulated killing. Nature. 2004;428(6985):868-71.

27. Balagadde FK, Song H, Ozaki J, Collins CH, Barnet M, Arnold FH, et al. A synthetic Escherichia coli predator-prey ecosystem. Molecular Systems Biology. 2008;4:8.

28. Fuqua WC, Winans SC. A LUXR-LUXI TYPE REGULATORY SYSTEM ACTIVATES AGROBACTERIUM TI PLASMID CONJUGAL TRANSFER IN THE PRESENCE OF A PLANT TUMOR METABOLITE. Journal of Bacteriology. 1994;176(10):2796-806.

29. Engelberg-Kulka H, Glaser G. Addiction modules and programmed cell death and antideath in bacterial cultures. Annual Review of Microbiology. 1999;53:43-70.

30. Guzman LM, Belin D, Carson MJ, Beckwith J. TIGHT REGULATION, MODULATION, AND HIGH-LEVEL EXPRESSION BY VECTORS CONTAINING THE ARABINOSE P-BAD PROMOTER. Journal of Bacteriology. 1995;177(14):4121-30.

31. Goyal G, Tsai SL, Madan B, DaSilva NA, Chen W. Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome. Microbial Cell Factories. 2011;10:8.

32. Hu B, Du J, Zou RY, Yuan YJ. An Environment-Sensitive Synthetic Microbial Ecosystem. Plos One. 2010;5(5):9.

33. Biliouris K, Babson D, Schmidt-Dannert C, Kaznessis YN. Stochastic simulations of a synthetic bacteria-yeast ecosystem. Bmc Systems Biology. 2012;6:13.

34. Brambilla M, Ferrante E, Birattari M, Dorigo M. Swarm robotics: a review from the swarm engineering perspective. Swarm Intelligence

Controlling Meiotic Recombination in Plants with RNA Guided Cas9

ABSTRACT

In sexually reproducing species, meiotic recombination produces novel allele combinations, generating genetic variation in offspring. Plant breeding relies on this variation for crop improvement. However, recombination events are stochastic and recombination rates between certain genes are often very low. The ability to manipulate the rate and location of recombination events within plant genomes would therefore provide a great breakthrough for crop improvement.

Meiotic recombination begins with the formation of double stranded breaks (DSBs) in aligned chromosomes catalysed by the highly-conserved protein Spo11. Plants with inactivating mutations in Spo11 have greatly reduced fertility. The endonuclease Cas9 combined with a single guide (sg)RNA can produce targeted DSBs in plants; therefore, it presents an attractive potential means to manipulate meiotic recombination, by reproducing the activity of Spo11.

The long-term goal of this study was to test the hypothesis that: targeted DSBs in aligned homologous chromosomes during meiosis, catalysed by a Cas9:sgRNA complex, can trigger recombination at a desired locus. This could be tested by transforming *Arabidopsis thaliana* Spo11 mutant plants with a construct containing Cas9, driven by a meiocyte-active promoter, and a sgRNA targeting multiple sites throughout the genome. In theory, the activity of Spo11 could be artificially replaced and evidenced by the recovery of fertility in T_1 plants homozygous for a Spo11 mutation. This would provide a proof of concept that could be applied to crops for the advancement of crop breeding.

As a first step towards this goal, a Cas9:sgRNA expressing binary vector was assembled and a published edit was replicated in *Arabidopsis,* to confirm that the construct was able to generate DSBs. Editing results were disappointing and are presented and discussed here. Also, Spo11 mutant *Arabidopsis* lines were acquired and mutations verified.

BACKGROUND

In sexually reproducing species, meiotic recombination produces novel allele combinations, generating genetic variation. Plant breeding relies on this genetic diversity for crop improvement: crosses produce offspring with new combinations of beneficial alleles and useful genes from wild relatives can be introgressed into high yielding varieties, to introduce traits such as pest resistance or draught tolerance (1). However, recombination events are stochastic and recombination rates between certain genes are often very low — because genes are in close physical proximity to one another on the chromosome (linkage drag) or in cold spots on chromosomes, where very little recombination occurs (2) —which limits the speed and efficiency of plant breeding. Although crop breeding has been greatly accelerated by advances in technologies such as marker assisted selection, ultimately it still relies on natural meiotic recombination to shuffle DNA, with all its associated infrequency and stochasticity. The ability to manipulate the rate and location of recombination events within plant genomes would therefore provide a great breakthrough for crop improvement (3). Meiotic recombination begins with formation of a double stranded break (DSB)

in a chromosome catalysed by the protein Spo11. The CRIPSR/Cas9 system can produce targeted DSBs in plants; therefore, it presents an attractive potential means to manipulate meiotic recombination.

Meiotic Recombination

Meiosis consists of a single round of DNA replication followed by two rounds of nuclear and cellular division, producing four gametes with half the ploidy number of the parent cell. Meiosis 1 segregates homologous chromosomes, meiosis 2 then segregates sister chromatids and fertilisation restores ploidy in the zygote to the parental number. Meiotic recombination takes place during prophase of meiosis 1, facilitated by chromosomal crossover — the reciprocal exchange of corresponding sections of DNA between homologous chromosomes — which is physically manifested by the appearance of chiasmata between homologues aligned at synapsis (4). Meiotic recombination is controlled by a complex set of interacting proteins, initiated by the highly-conserved protein Spo11 catalysing a DSB in one of the homologues (5–8) (**Figure 1**). Only a subset of these breaks will be resolved as cross over events (9). A suite of proteins then orchestrates: resectioning of DNA near the DSB, leaving single stranded DNA; strand invasion of the homologous sequence; templated repair; and finally cutting to release the recombined homologues (1). Chromosomal crossover facilitates synapsis and chromosomal alignment at the meiocyte equator during meiosis 1, and is therefore necessary for the correct segregation of homologues into daughter cells (10). An absence of crossovers results in aneuploid daughter cells and reduced fertility (11–13).

In *Arabidopsis thaliana*, megasporocytes undergo meiosis inside the ovary, with only one of the four daughter cells surviving as a single megaspore which develops into an embryo sac — the female gametophyte. Within the anther, microsporocytes undergo meiosis producing four microspores, each of which develops into a pollen grain — the male gametophyte (14). In *Arabidopsis* there are three Spo11 homologues, two of these, AtSPO11-1 and AtSPO11-2, are believed to function in the formation DSBs during recombination (11,12,15–18). Plants homozygous for a knock out mutation in either *Spo11-1* or *Spo11-2* have short siliques and have greatly reduced fertility; furthermore, plants heterozygous for a knockout at both *Spo11-1* and *Spo11-2* show a similar loss of fertility (11–13,18). Spo11 mutants therefore present a potential test system for artificially induced meiotic recombination. If recombination events could be triggered in plants homozygous Spo11 mutations, this could theoretically recover fertility.

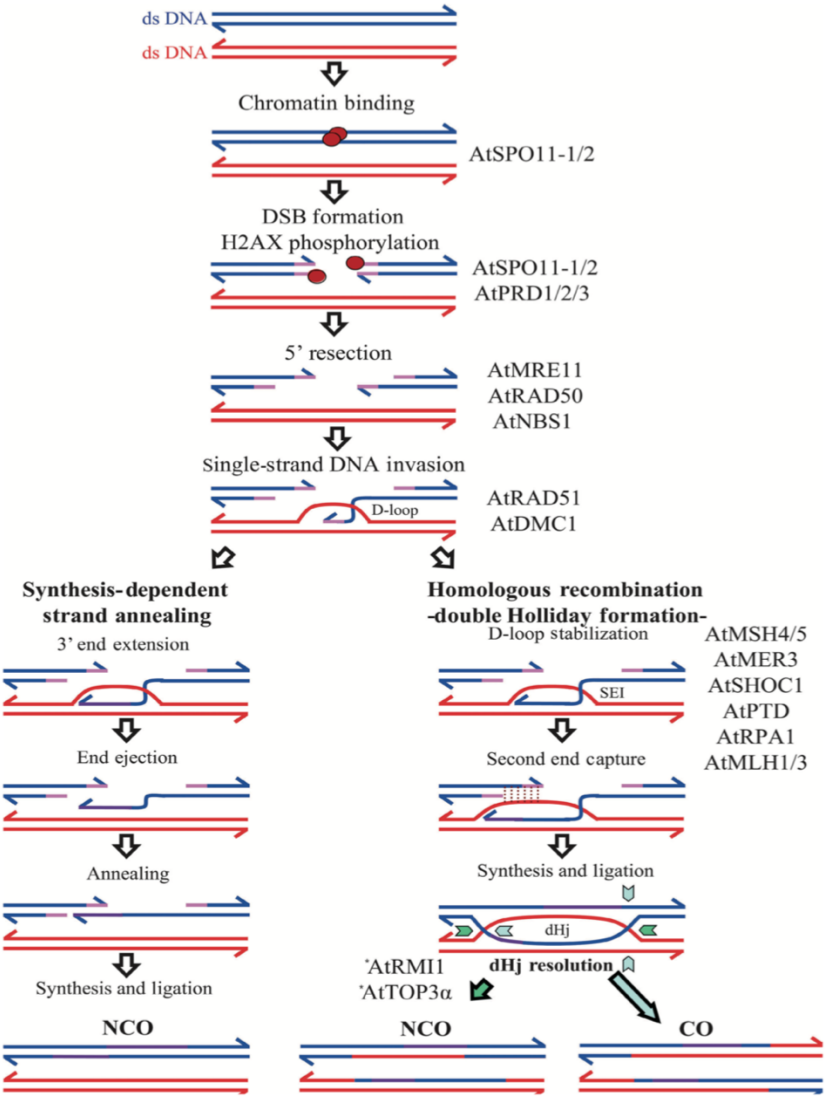

Figure 1. Key stages of meiotic recombination as elucidated in *A. thaliana*. Showing, the potential non-crossover (NCO) and crossover (CO) outcomes of the process. Proteins known to be involved at each stage are listed on the right; note AtSPo11-1/2 at double strand break (DSB) formation. Image taken from (10).

The CRISPR/Cas9 System

The RNA guided Cas9 system has been demonstrated as an efficient method to create targeted genome edits in plants (19–25).The system comprises of two components: the Cas9 nuclease, which catalyses the formation of DSBs in DNA and the single guide RNA (sgRNA), which provides target specificity by the binding of a 20 nucleotide 'guide sequence', at its 5' end, with a complementary 'target sequence' in the genomic DNA, while a 'scaffold sequence', at its 3' end, forms a secondary structure capable of recruiting the Cas9 protein to the target (26) (**Figure 2**). The Cas9 protein contains two nuclease domains: HNH and RuvC which cleave the DNA target- and non-target- strands respectively.

The system is derived from the CRISPR:Cas9 adaptive immune system of *Streptococcus pyogenes* and much of the terminology used can be confusing due to parallels between the functioning of the natural system and of the artificial genome editing system (27,28). In the natural system of bacteria, DNA regions termed clustered regularly interspaced short palindromic repeat (CRISPR) arrays consist of partially palindromic repeats interspersed with unique 'spacer' sequences. It was discovered that a short sequence of invading viral DNA termed a 'protospacer' is excised by CRISPR associated (Cas) proteins, Cas1 and Cas2, and incorporated into the CRISPR array to create a new spacer. Upon a future viral infection, the CRISPR array region is transcribed into an RNA which is then processed by other Cas proteins into a CRISPR (cr)RNA — derived from the spacer DNA — and a trans-encoded CRISPR (tracr)RNA — derived from the CRISPR DNA sequence. The crRNA and tracrRNA join via a short complementary sequence to form a structure able to recruit Cas9 to cleave the invading DNA (27,28). When the system was appropriated for genome editing, it was found it could be simplified by expressing the crRNA and tracrRNA as a single guide (sg)RNA (26). Therefore, the guide, scaffold and target sequences referred to in the editing literature can be thought of as equivalent to the spacer/crRNA, CRIPSR/tracrRNA and protospacer sequences respectively, of the natural system. In bacteria, to prevent Cas9 cleaving the host genome at the CRISPR array, a DNA binding domain of Cas9 needs to recognise a 3 base pair sequence directly adjacent to the protospacer sequence, termed the 'protospacer adjacent motif' (PAM) which is not present in the CRIPSR array. The canonical PAM of *S. pyogenes* Cas9 is NGG and the presence of this sequence directly following the selected target sequence is the only prerequisite when choosing a target editing site (26).

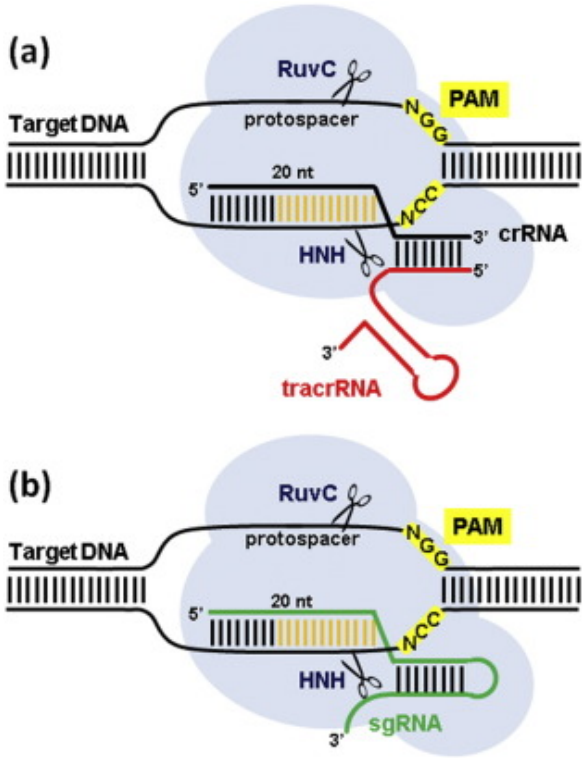

Figure 2. Components of the CRISPR/Cas 9 system. crRNA, (a) Shows separate CRISPR (cr)RNA and trans-encoded CRISPR (tracr)RNA as in the natural system. (b) Shows the RNA components combined as a single guide (sg)RNA as used the majority of genome editing studies. The Cas9 protein is shown in grey with the HNH and RuvC nuclease domains cleaving the two DNA strands. PAM = Protospacer Adjacent Motif. Image taken from (23).

sgRNA:Cas9 Editing in Plants

Arabidopsis comes second to rice (*Oryza sativa*), in terms of the greatest number of reported edits for a plant species (29). In *Arabidopsis*, RNA guided Cas9 has been used to produce edits in protoplasts, and to create stable heritable edits in whole plants (19,21–24,30–41). Most studies have been performed with the sgRNA and Cas9 sequences in the same T-DNA construct, which is integrated into the plant genome via *Agrobacterium* mediated transformation using the floral dip method (42). The most popular promoter choice for Cas9 has been CaMV 35S or double CaMV 35S, and sgRNAs are usually expressed from *Arabidopsis* U6 small nuclear RNA (snRNA) promoters, which are transcribed by RNA polymerase III (23,43–45). The most frequently performed type of editing relies on the endogenous DNA repair pathway of Non-Homologous End Joining (NHEJ) which is error prone and so will often result in short insertions/deletions (indels) at the site of a single DSB created by Cas9. This technique is commonly used to create 'knock outs' via frame shifts and genes or alternatively to delete a longer section of DNA by making two flanking cuts (21,25,31,38,46–48). A second method utilizes the Homology Directed Repair (HDR) pathway to introduce a desired sequence into the genome (Gene Targeting, GT); to achieve this, a template with flanking regions homologous to genomic sequences either side of the cut site needs to be introduced to the cell, in addition to the Cas9:sgRNA construct. While this method offers a greater flexibility of possible genome alterations, it is much harder to apply, and the proportion of plants showing the alteration is far lower than with NHEJ(49).

Project Aims

The long-term goal of this study was to test was the hypothesis: that by transforming *Arabidopsis* Spo11 mutants with a Cas9:sgRNA construct expressed during meiosis, the activity of Spo11 could be replaced and fertility would be recovered. A published Cas9:sgRNA construct would be modified to have Cas9 driven by a meiocyte-active promoter and with a sgRNA targeting a sequence present at multiple sites throughout the genome. T_0 plants heterozygous for a *Spo11-1* or *Spo11-2* knock out mutations were to be transformed using *Agrobacterium* mediated transformation. DSBs created in aligned homologous chromosomes in meiocytes of Spo11 -/- T_1 plants could theoretically trigger recombination and recover fertility.

As a first step towards this long-term goal, a Cas9:sgRNA expressing binary vector was obtained and an attempt made to replicate an edit from the original publication (47) in *Arabidopsis,* to confirm that the construct was able to generate DSBs. The editing ability of the construct was tested at a single-locus, replicating as closely as possible the published methods. Had the editing ability been verified, the construct promoter would have been swapped for a meiotically active promoter and the sgRNA would have been changed to target a sequence present in all chromosomes: in order to move on to the aspirational recombination experiments in Spo11 mutants. Spo11 mutant lines were also acquired and mutations verified, in preparation for the aspirational experiments.

MATERIALS AND METHODS

Selection of Cas9:sgRNA Vector

A number of Cas9:sgRNA binary vectors have been created for Arabidopsis; these differ in terms of the species Cas9 is codon optimized for, choice of promoters driving Cas9- and sgRNA expression and choice of binary vector backbone (29). For this study, a construct created by Ma *et al.* (47) was selected based on the following criteria: the construct had been verified for editing ability in *Arabidopsis*; gene editing efficiency was reportedly high; details of the methodology used are clear from the publication; and the construct was designed to allow straight forward insertion of a new guide sequence into the sgRNA cassette. To confirm that the published editing results were repeatable, an editing experiment from the original paper was replicated, keeping all steps as close to those reported as possible. The two plasmids needed to assemble a binary vector carrying Cas9 and sgRNA were purchased from Addgene (catalogue numbers 66191 and 66203), these are pYLCRISPR/Cas9P35S-N and pYLsgRNA-AtU6-29, and are based on pUC18 and pCAMBIA1300 backbones respectively. pYLsgRNA-AtU6-29 is an intermediate vector carrying a sgRNA expression cassette, and pYLCRISPR/Cas9P35S-N is a binary vector that carries, between T-DNA borders: the Cas9 gene (codon optimised for rice and expressed from a double CaMV 35S promoter, with nuclear localisation sequences at each end); the plant selectable marker gene neomycin phosphotransferase II; and an insertion site for the sgRNA expression cassette (between two BsaI restriction sites, that flank a toxic ccdB gene which acts as a negative selectable marker).

Selecting a Target Sequence

The target sequence previously selected by Ma *et al.* is in gene At5g55580, which encodes a putative mitochondrial transcription termination factor protein (mTERF), of unknown function. Sequence: GATGAGTTCGAGGAAGTATG(TGG) — the three bases in brackets are the PAM sequence.

Ma *et al.* reported that 3 out of 9 T_1 plants, carrying the T-DNA insert and screened at this locus, showed evidence of editing events (2 heterozygous and 1 biallelic), and that some plants with mutations in this gene showed albino phenotypes, from which they concluded that the gene is essential for chloroplast biogenesis.

Plasmid Assembly

For full details of assembly steps see Ma *et al.* (47). Briefly: two oligos corresponding to the two strands of the 20 base pair guide sequence flanked by BsaI restriction sites were synthesized by Eurofins (At-1F: ATTGATGAGTTCGAGGAAGTATG and At-1R: AAACCATACTTCCTCGAACTCAT); these were annealed and the resulting guide sequence was cloned into plasmid pYLsgRNA-AtU6-29 to give a functional sgRNA expression cassette. A nested PCR was performed with this modified pYLsgRNA-AtU6-29 as a template, using two sets of flanking primers to create an amplicon of the complete sgRNA expression cassette

(1st PCR: U-F: CTCCGTTTTACCTGTGGAATCG and gR-R: CGGAGGAAAATTCCATCCAC;

2nd PCR Pps-GGL: TTCAGAGGTCTCTCTCGACTAGTATGGAATCGGCAGCAAAGG and Pgs-GGR: AGCGTGGGTCTCGACCGACGCGTATCCATCCACTCCAAGCTC).

The primers of the second PCR round introduced BsaI restriction sites that gave sticky ends compatible with insertion into the vector pYLCRISPR/Cas9P35S-N. Finally, the PCR amplicon was cloned into pY-LCRISPR/Cas9P35S-N to replace the ccdB gene and create a Cas9 : sgRNA binary vector, hereafter referred to as **Construct 1** (**Figure 3**). Construct 1 was verified by Sanger sequencing.

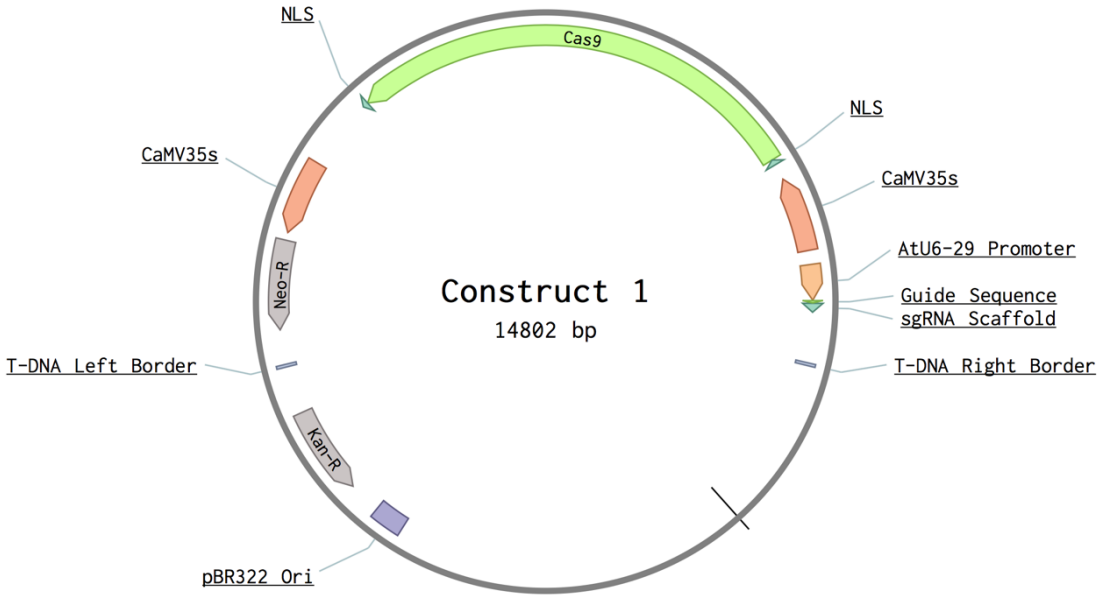

Figure 3. Plasmid map of Construct 1 – used to create an edit in gene At5g55580. NLS = Nuclear localisation sequence (SV40 at start: nucleoplasmin at end), Ori = Origin of replication, Kan-R = Kanamycin resistance gene, Neo-R = Neomycin/G418 resistance gene (Neomycin phosphotransferase II).

Growing Plants

All plants used were Col-0 ecotype. For floral dipping, plants were grown as "muffins" in 3" pots in a 3:1 mix of compost : sand, and grown in greenhouse conditions at 21°C with 16 hours light, 8 hours dark cycles under cool, white fluorescent bulbs at a light intensity of approximately 100 mmol m^{-2} s^{-1}. For all other experiments, plants were grown in 4, 4, 5 cm trays under the same conditions. All seeds were stratified for 48 h in darkness at 4°C prior to planting.

Agrobacterium Transformation

The complete Cas9:sgRNA construct (Construct 1) was introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation, and transformation of six week old flowering *Arabidopsis* was performed as described previously (42). 25 plants were treated (5 muffins x 5 plants per muffin). Plants in a sixth untreated muffin were used as controls.

Selection of Transformed Plants

T_1 seeds were surface sterilized by submerging in 70% ethanol for 5 minutes and rinsing in 3 washes of distilled water, and then plated on 1% agar plates supplemented with 0.5% Murashige and Skoog Growth Medium and 25µg

mL^{-1} G418 for selection, approximately 1000 seeds per 90mm diameter plate x 5 plates (16 seeds cm^{-2}). Plates were stratified and kept in a controlled growth room at 21°C with 16 hours light, 8 hours dark cycles under cool, white fluorescent bulbs at a light intensity of approximately 100 mmol m^{-2} s^{-1}. After 10 days, healthy looking seedlings were transferred to soil and grown as above.

PCR Screening Transformed Plants

An approximately 0.5 cm diameter section of one of the first true leaves was taken from each plant for gDNA extraction. Integration of the T-DNA was confirmed by performing a PCR with primers specific to the Cas9 gene (pYLCas9-3F: GATGGACTGGTGGATGAGAGT and pYLCas9-4R CATCGACAACAAGGTCCTCAC) to produce a 1504 base pair amplicon.

Sequencing to Detect Edits

PCR amplifications were carried out using primer pairs flanking the At5g55580 target site (Ed1P1-F: CTTTAGGATGATTTCGCTGAGG and EdP1-R GGAATTAGGAATGCCAAGACCC) to produce a 689 base pair amplicon. The PCR products were Sanger sequenced using the same primers. Evidence of an editing event can be detected by analysing the sequenced chromatograms by eye; the sudden appearance of double peaks at the target sequence, with a reduction in quality scores for the following sequence indicates a potential editing event (47).

Verification of Spo11-1 and Spo11-2 Mutants

Spo11-1 and *Spo11-2* mutants were purchased from the *Arabidopsis* stock centre (Lines SALK_146172 and GABI 749C12). To bulk seed and verify the correct mutations were present, plants were grown and screened by performing a PCR with primers specific to the T-DNA insert and relevant *Spo11* locus:-

Primers used were: for *Spo11-1* wild type amplicon SALK172-F: TTTCAGTGTAGTCGGTACAACTTGAATG-TG and SALK172R: CCACAACCAGTATGTACTCAGCTAAGCTAAC; for *Spo11-1* T-DNA insertion amplicon SALKLBb1: GCGTGGACCGCTTGCTGCAACT and SALK172R; for *Spo11-2* wild type amplicon P1: TCCTGATCTGCCAATTCTTG and P2: CCATGACAATAGAGAGCTTC; for *Spo11-1* T-DNA insertion amplicon P1 and GABILB: TTGGACGTGAATGTAGACAC.

RESULTS

40 T_1 plants were grown: 36 putative transformants from selection plates and 4 control plants (the offspring of plants grown under the same conditions but not inoculated with *Agrobacterium*). The plants were screened for T-DNA integration by PCR with Cas9 specific primers, and all except two of the 36 putative transformants produced an amplicon of the correct size (**Figure 4**), the two plants with no amplicon could be plants without the T-DNA insert that somehow survived selection, or the lack of an amplicon could simply be down to a failure at some stage of the gDNA extraction or PCR preparation, the two plants were omitted from further analysis. This gives a successful transformation rate of 0.8%, which is consistent with published results (42). The 34 positive plants were Sanger sequenced for a region including the target editing site; only one had detectable evidence of an editing event. The chromatogram of plant 33 had visible but weak secondary peaks appearing from 4 bases before the PAM sequence; 3 base pairs before the PAM is documented as having the highest probability of a mutation (19,20,22,24,25,40,44,50). Inspection shows that the second set of peaks is consistent with a single base pair insertion probably an A (**Figure 5**).

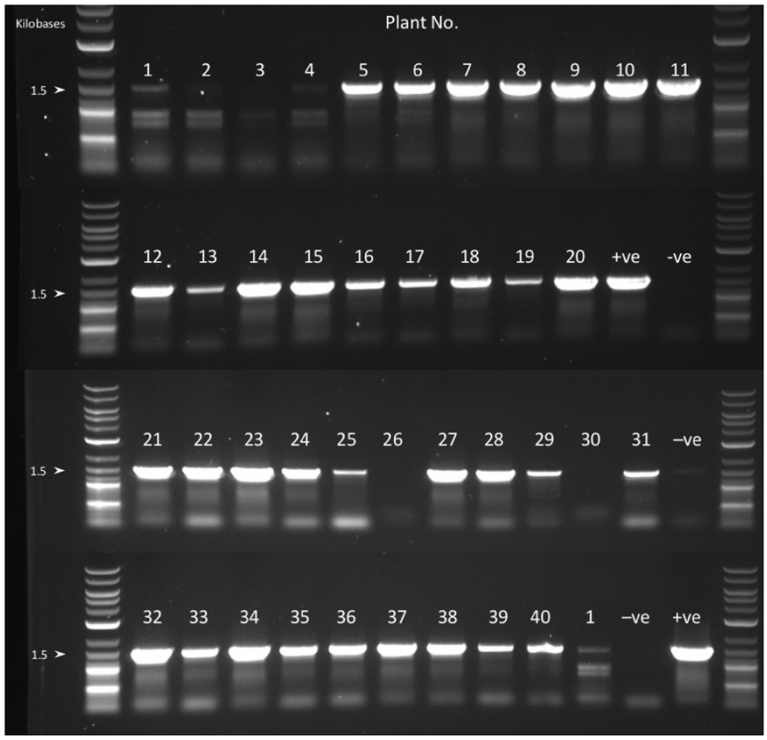

Figure 4. Gel of the products of PCR for putative transformants. Performed with primers specific to the Cas9 gene (primers pYLCas9- 3F and 4R) from the gDNA of putative transformants; with the exceptions of plants 1-4 which are *Agrobacterium* unexposed controls, and plants 26 and 30, all plants have an amplicon of the expected size (1504 bp), indicating successful integration of the Cas9:sgRNA construct. –ve = water control, +ve = original Construct 1 plasmid DNA used as template.

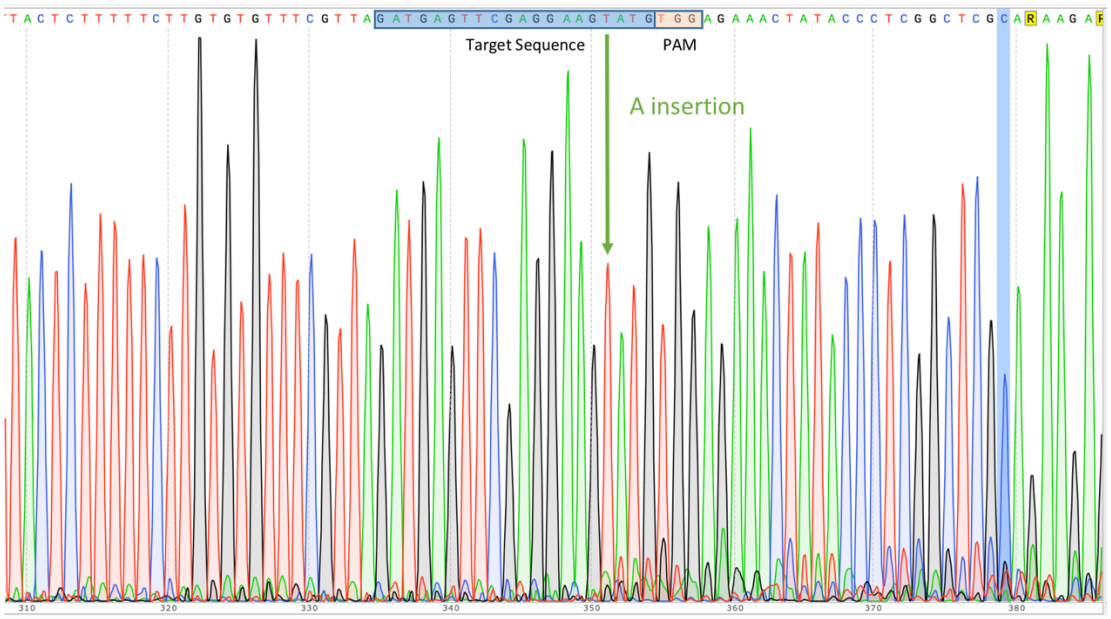

Figure 5. Chromatogram of the putative editing site of plant 33. Produced by sequencing the genomic DNA of plant 33 at the target editing site in gene At5g55580. An extra set of peaks appear from four base pairs before the PAM sequence and this new sequence is consistent with an insertion of an extra A.

Original sequence: GATGAGTTCGAGGAAGTATGTGGAGAAACTATACCCTCGGCTCGC

New sequence: GATGAGTTCGAGGAAGATATGTGGAGAAACTATACCCTCGGCTCGC

Plasmids with Meiotically Active Promoters

Had the single-locus editing experiment produced satisfactory results, the next step toward the long-term experimental goal of generating multiple DSBs in meiocytes — to test the hypothesis that recombination can be artificially triggered — would require modifications to Construct 1. To create DSBs during meiosis, Construct 1 needed to be modified to have Cas9 expressed from a meiocyte-active promoter and to include sgRNAs which target a sequence that occurs 500-1000 times throughout the *Arabidopsis* genome (this matches estimates of the number of crossovers occurring in *Arabidopsis* (52)). In total, three constructs were designed and partially assembled, with three different promoters shown to be preferentially expressed in meiocytes. These promoters were selected from a study by Li *et al.* (53) and were first identified as meiotically active using transcriptome analysis and then verified by cloning into a GFP reporter construct and checking for fluorescence in meiocytes. The promoters are:

1. PMS5, The MS5 gene is expressed throughout meiosis and is essential for male meiosis.
2. PAT1G15320 expression is limited to later meiosis
3. PAT4G40020 expression is limited to early meiosis

Therefore, PAT4G40020 is the promoter most likely to drive Cas9 expression at the correct time and to rescue fertility. Of course, an obvious way to ensure Cas9 expression is at the correct time would be to use the endogenous Spo11 promoter, but this has its own inherent problems in terms of little information about its functioning being known, and the potential of problems in compatibility with the Cas9 construct.

Plasmid Assembly

Promoter regions were PCR amplified from Col-0 genomic DNA, using primers designed by Li *et al.* (53) modified to add restriction sites compatible with cloning into Construct 1 or pYLCRISPR/Cas9P35S-N. The PMS5 PCR product was cloned into Construct 1 to make **pMS5-Construct1** (**Figure 6**), and this construct has been sequenced. However, due to non-specific amplification and problems experienced with gel extraction, and time limitations, the other plasmids were not assembled.

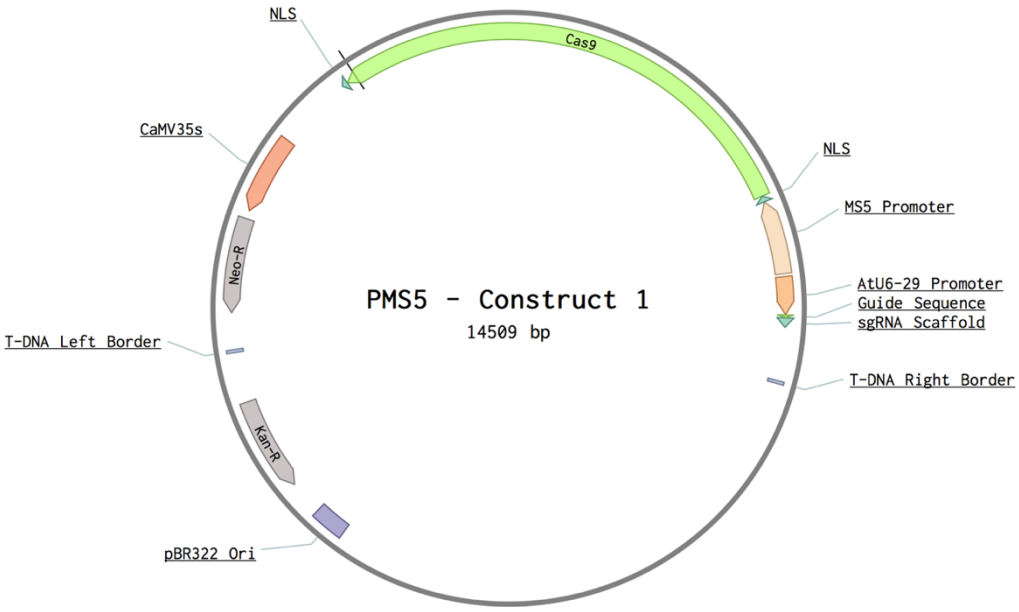

Figure 6. Plasmid map of pMS5-Construct 1 – a modified version of Construct 1 with the promoter sequence of the MS5 gene replacing the CaMV 35S promoter. NLS = Nuclear localisation sequence (SV40 at start: nucleoplasmin at end), Ori = Origin of replication, Kan-R = Kanamycin resistance gene, Neo-R = Neomycin/G418 resistance gene (Neomycin phosphotransferase II).

Verification of Spo11-1 and Spo11-2 Mutants

Spo11-1 and *Spo11-2* mutant lines – SALK_146172 and GABI 749C12 – were grown and screened by performing a PCR with primers specific to the T-DNA insert and relevant *Spo11* locus. All results indicated lines were segregating for T-DNA insertions at the loci as expected.

DISSCUSION

The evidence of editing in plant 33 supports the ability of the construct to create DSBs; however, our editing efficiency was far lower than that reported by Ma *et al.* (47) - 1/34 plants compared to 3/9 plants. The methods used here replicated those of the original publication, so it unclear why there is such a large discrepancy. Also, the secondary peaks of the chromatogram of plant 33 are much smaller than those illustrated by Ma *et al.*, and this would indicate chimeric plants with only a small proportion of total cells deriving from an editing event. Ma *et al.* reported heterozygous or biallelic mutations which would give two sets of peaks with comparable heights. Other reports suggest chimeric T_1 plants are the most common outcome of Cas9 editing *Arabidopsis* using similar methodology, and the methods employed to detect edits usually take account of this by using PCR based detection methods that amplify edited sequences amongst a background of none edited cells (25,31,34,46,50,51). This begs the question of why Ma *et al* produced such efficient results, one explanation would be that some detail of their methods encouraged edits at a very early stage of T_1 plant development.

The results suggest that Construct 1 is capable of creating gene editing events in *Arabidopsis* that can be detected in T_1 plants, but that the construct is not very efficient and that edits will have a low probability of being heritable. Plasmid assembly was easily achieved and the whole plasmid was sequence verified. The floral dip was also successful, evidenced by a high proportion of transgenic T_1 plants verified by antibiotic selection and PCR for the T-DNA. Therefore, the problems would appear to be in expression of Cas9 or sgRNA within transformed cells, or the functioning of the editing complex.

Design Weaknesses

In light of recently published papers on Cas9 editing *Arabidopsis*, several weaknesses of our experimental design become apparent.

The first problem is the choice of promoter. As mentioned earlier, Cas9 editing studies in *Arabidopsis* have often produced chimeric T_1 plants; this is probably because the widely used CaMV35S promoter is only active in later stages of plant development. The CaMV35 promoter is not known to be active during early development and therefore Cas9 is expressed too late to generate edits that will pass to all cells. Several papers have reported this finding and improvements with promoters active in early development. Examples are the EC1 promoter (38) and pAT4G40020 (53). Studies have shown that by expressing Cas9 from a promoter active in earlier development editing appears in a greater proportion of T_1 plants and homozygous mutations are far more frequent (38,40,53). Therefore, an interesting side line to this study would be to test the meiosis promoter constructs on the target used for Construct 1. Theoretically, any edit made during meiosis should be passed to all cells, and this could be much more efficient than the CaMV35S driven construct for general editing purposes.

Two plasmid designs, with different promoters driving Cas9 expression, have now emerged as the most frequently used to edit *Arabidopsis*: 1) pHEE401 using the EC.1.2 promoter from Wang *el al.* (38). The authors trialled the EC1.2 promoter in combination with different terminators and enhancers, and achieved around 10% triple homozygous or biallelic mutants in T_1 plants, with the best combination. 2) pEn-Chimera using promoter pDE-pUbi-Cas9 from Fauser *et al.* (35). The authors tested the potential to edit *Arabidopsis* with Cas9 as nuclease and with a modified Cas9 as a nickase. Using the nuclease, approximately one third of reads from deep sequencing were edited. Future plasmid designs should probably utilise these plasmid designs as a base.

The second weakness in our plasmid design was that it included only one sgRNA sequence; most published studies have used two sgRNAs, targeting sequences in close proximity in the target region. It is easy to see how this redundancy gives improved editing efficiency: firstly it is known that editing efficiency is highly variable between sgRNAs (20); therefore, two sgRNAs gives a greater chance that one of sequences will be efficient. And secondly it is known that the majority of repairs of DSBs are perfect repairs, and so two sgRNAs give the additional chance that two synchronous DSBs will result in a short deletion (25,34,40,46,51).

Another issue in the experimental design, that could account for our findings, is the method of screening: Sanger (direct) sequencing is likely to miss many chimeric plants if edited cells are greatly outnumbered by non-edited cells. Indeed, for plant 33 the secondary peaks in the chromatogram are small in comparison to the wildtype peaks, showing that the editing could have been easily overlooked. Screening for edits in this way is not only prone to miss chimeric editing but is also time consuming, leading to a small number of plants being screened. Many studies get around these problems by pooling DNA from batches of plants and then either deep sequencing or using assays based on selective restriction digests at the target site followed by PCR. Both methods are more sensitive to chimeric plants and can be used to screen large numbers of plants. Positive batches can then by subdivided, and edited plants homed in on. Future work should use a combination of these screening methods. Additionally, targeting of a gene that will give a visual phenotype change is useful for ease of screening. Our study focused on an edit with a phenotypic effect but other target genes have been utilised in more than one study and therefore may be more reliable. An often-used target is the PDS3 gene, as it gives albino dwarf phenotype and future work could move to this target as multiple reports of successful target sequences exist (31,35,37).

There are other potential explanations for the performance of the editing construct, such as: choice of sgRNA promoter, codon optimisation of Cas9 and choice of target sequence. However, none of these seem as likely a cause as the three issues mentioned above, based upon the importance of variables mentioned in the literature. A more controlled, but time consuming, protocol for the development of a new editing construct for *Arabidopis* could include an *in vitro* assay of sgRNA binding using restriction digest – PCR, Surveyor or T7 assays, and test of editing efficiency in protoplasts.

Future Work

Interesting results can still be gained from plant 33: seed has been harvested from the plant and a simple antibiotic screen could provide evidence of the heritability of the T-DNA. Any resistance would demonstrate that T-DNA had indeed stably integrated, and the ratio of resistant to non-resistant seedlings would give an indication of the number of insertions. Screening T_2 plants, from plant 33 and other T_1 plants, for editing events would give further evidence to hypotheses about the somatic nature of the editing achieved with Construct 1.

To date, CRISPR/Cas9 editing in *Arabidopsis* has not settled on any one set protocol or plasmid design, and there has been no side by side test of the most widely used plasmids, to our knowledge. Therefore, plasmid designs that stand out in the literature could be tested for editing efficiency under the same conditions and for the same target sequence. Assuming efficient germ line edits can be achieved with at least one of the plasmids, the three meiocyte active promoters mentioned above could replace the promoter in order to conduct the Spo11 rescue experiments.

Towards achieving our long-term goal, an alternative construct design would comprise an inactive 'dead' Cas9 (dCas9) fusion with the native SPO11 protein. There are several underlying assumptions in the hypothesis that DSBs created by Cas9 in meiocytes can recover fertility in Spo11 mutants — even overlooking the problems of achieving DSBs precisely during meiosis. SPO11 is known to have roles other than DSB creation; it also known to recruit other proteins involved recombination to the site of the break (11-13). Therefore, a fusion protein including SPO11 could circumvent this limitation. Also, it is known that most DSBs formed during meiosis do not result in recombination, but are repaired via other pathways (1,2,9,10). A SPO11-GAL4 fusion protein has already been demonstrated to increase meiotic recombination rate in yeast *Saccharomyces cerevisiae* (5).

Another consideration must be that given the difficulty of Cas9 editing in *Arabidopsis*, it may be more efficient to attempt gene targeting or manipulation of meiotic recombination rate in crop species directly. Certainly, the literature suggests that CRISPR/Cas9 editing can be achieved more efficiently in a number of crop plants including rice and barley (32,34).

Despite its difficulties, the goal of genome engineering plants using RNA-guided Cas9 has huge potential for both agriculture and fundamental research, and is an exciting line of research.

REFERENCES

1. Zhou, A. & Pawlowski, W. Regulation of meiotic gene expression in plants. *Front. Plant Sci.* **5,** (2014).

2. Salomé, P. *et al.* The recombination landscape in Arabidopsis thaliana F2 populations. *Hered.* **108,** 447–455 (2011).

3. Wijnker, E. & Jong, H. Managing meiotic recombination in plant breeding. *Trends in plant science* **13,** 640–646 (2008).

4. Schuermann, D., Molinier, J., Fritsch, O. & Hohn, B. The dual nature of homologous recombination in plants. *Trends Genetics* **21,** 172–181 (2005).

5. Murakami, H. & Nicolas, A. Locally, meiotic double-strand breaks targeted by Gal4BD-Spo11 occur at discrete sites with a sequence preference. *Mol. Cell. Biol.* **29,** 3500–16 (2009).

6. De Massy, B. Initiation of Meiotic Recombination: How and Where? Conservation and Specificities Among Eukaryotes. *Genetics* **47,** 563–599 (2013).

7. Bergerat, A. *et al.* An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386,** 414–7 (1997).

8. Hartung, F. *et al.* An Archaebacterial Topoisomerase Homolog Not Present in Other Eukaryotes Is Indispensable for Cell Proliferation of Plants. *Curr. Biology* **12,** (2002).

9. Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N. & Grelon, M. The Molecular Biology of Meiosis in Plants. *Annu. Rev. Plant Biology* **66,** 1–31 (2015).

10. Osman, K., Higgins, J., Sanchez-Moran, E., Armstrong, S. & Franklin, F. Pathways to meiotic recombination in Arabidopsis thaliana. *Phytologist* **190,** 523–544 (2011).

11. Stacey, N. *et al.* Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J.* **48,** 206–216 (2006).

12. Grelon, M., Vezon, D., Gendrot, G. & Pelletier, G. AtSPO11-1 is necessary for efficient meiotic recombination in plants. *Embo J* **20,** 589–600 (2001).

13. Sanchez-Moran, E., Santos, J.-L., Jones, G. & Franklin, F. ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in Arabidopsis. *Genes Dev.* **21,** 2220–2233 (2007).

14. Li, J., Hsia, A.-P. & Schnable, P. Recent advances in plant recombination. *Current Opinion in Plant Biology* **10,** 131–135 (2007).

15. Hartung, F. & Puchta, H. Molecular characterisation of two paralogous SPO11 homologues in Arabidopsis thaliana. *Nucleic Acids Res.* **28,** 1548–1554 (2000).

16. Hartung, F. & Puchta, H. Molecular characterization of homologues of both subunits A (SPO11) and B of the archaebacterial topoisomerase 6 in plants. *Gene* **271,** 81–86 (2001).

17. Shingu, Y., Mikawa, T., Onuma, M., Hirayama, T. & Shibata, T. A DNA-binding surface of SPO11-1, an Arabidopsis SPO11 orthologue required for normal meiosis. *FEBS J.* **277,** 2360–74 (2010).

18. Hartung, F. *et al.* The Catalytically Active Tyrosine Residues of Both SPO11-1 and SPO11-2 Are Required for Meiotic Double-Strand Break Induction in Arabidopsis. *Plant Cell Online* **19,** 3090–3099 (2007).

19. Shan, Q. *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31,** 686–8 (2013).

20. Mao, Y. *et al.* Application of the CRISPR–Cas System for Efficient Genome Engineering in Plants. *Mol. Plant* **6,** 2008–2011 (2013).

21. Lowder, L., Zhang, D., Baltes, N. & Paul, J. A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiology* **169,** 971–985 (2015).

22. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Sci.* **339,** 819–823 (2013).

23. Bortesi, L. & Fischer, R. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology Adv.* **33,** (2014).

24. Belhaj, K., Chaparro-Garcia, A., Kamoun, S. & Nekrasov, V. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* **9,** (2013).

25. Liu, W. *et al.* A detailed procedure for CRISPR/Cas9-mediated gene editing in Arabidopsis thaliana. *Sci. Bulletin* **60,** (2015).

26. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Sci.* **337,** 816–821 (2012).

27. Marraffini, L. CRISPR-Cas immunity in prokaryotes. *Nature* **526,** 55–61 (2015).

28. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.* **82,** 237–66 (2013).

29. III, J. W. & Qi, Y. CRISPR/Cas9 for plant genome editing: accomplishments, problems and prospects. *Plant cell reports* 1–11 (2016). doi:10.1007/s00299-016-1985-z

30. Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. & Kamoun, S. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotechnology* **31,** 691–693 (2013).

31. Li, J.-F. *et al.* Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat. Biotechnology* **31,** (2013).

32. Lawrenson, T, Shorinola, O, Stacey, N & Li, C. Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. *Genome Biology* (2015). doi:10.1186/s13059-015-0826-7

33. Ji, X., Zhang, H., Zhang, Y., Wang, Y. & Gao, C. Establishing a CRISPR–Cas-like immune system conferring DNA virus resistance in plants. *Nat Plants* **1,** (2015).

34. Jiang, W. *et al.* Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res* **41,** e188–e188 (2013).

35. Fauser, F., Schiml, S. & Puchta, H. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in Arabidopsis thaliana. *Plant J.* **79,** (2014).

36. Woo, J. *et al.* DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat Biotechnol* **33,** (2015).

37. Schiml, S., Fauser, F. & Puchta, H. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. *Plant J.* **80,** (2014).

38. Wang, ZP, Xing, HL, Dong, L, Zhang, HY & Han, CY. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome biology* (2015). doi:10.1186/s13059-015-0715-0

39. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences* **112,** 3570–3575 (2015).

40. Mao, Y. *et al.* Development of germ-line-specific CRISPR-Cas9 systems to improve the production of heritable gene modifications in Arabidopsis. *Plant Biotechnol J* (2015). doi:10.1111/pbi.12468

41. Yan, W, Chen, D & Kaufmann, K. Efficient multiplex mutagenesis by RNA-guided Cas9 and its use in the characterization of regulatory elements in the AGAMOUS gene. *Plant Methods* (2016). at <http://plantmethods.biomedcentral.com/articles/10.1186/s13007-016-0125-7>

42. Clough, S. & Bent, A. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J Cell Mol Biology* **16,** 735–43 (1998).

43. Osakabe, Y & Osakabe, K. Genome editing with engineered nucleases in plants. *Plant and Cell Physiology* (2015). doi:10.1093/pcp/pcu170

44. Upadhyay, S., Kumar, J., Alok, A. & Tuli, R. RNA-guided genome editing for target gene mutations in wheat. *G3* **3,** 2233–2238 (2013).

45. Schaeffer, S. & Nakata, P. CRISPR/Cas9-mediated genome editing and gene replacement in plants: Transitioning from lab to field. *Plant Sci* **240,** 130–142 (2015).

46. Feng, Z. *et al.* Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* **23,** 1229–32 (2013).

47. Ma, X. *et al.* A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular plant* **8,** 1274–1284 (2015).

48. Xing, H.-L. *et al.* A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC plant biology* **14,** 327 (2014).

49. Zhao, Y. *et al.* An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Scientific Reports* (2016). doi:10.1038/srep23890

50. Feng, Z. *et al.* Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc Natl Acad Sci* **111,** 4632–4637 (2014).

51. Jiang, W., Yang, B. & Weeks, D. P. Efficient CRISPR/Cas9-mediated gene editing in Arabidopsis thaliana and inheritance of modified genes in the T2 and T3 generations. *PLoS ONE* (2014). doi:10.1371/journal.pone.0099225

52. Giraut, L. *et al.* Genome-wide crossover distribution in Arabidopsis thaliana meiosis reveals sex-specific patterns along chromosomes. *PLoS Genet.* **7,** e1002354 (2011).

53. Li, J. *et al.* Characterization of a set of novel meiotically-active promoters in Arabidopsis. *BMC plant biology* **12,** 1 (2012).